Stochastic Environmental Research and Risk Assessment (2018) 32:1531–1544
https://doi.org/10.1007/s00477-018-1547-8

ORIGINAL PAPER

Using a Bayesian belief network model for early warning of death and severe risk of HFMD in Hunan province, China

Yilan Liao[1,2] · Bing Xu[1,3,6] · Xiaochi Liu[1,4] · Jinfeng Wang[1,2] · Shixiong Hu[5] · Wei Huang[5] · Kaiwei Luo[5] · Lidong Gao[5]

Published online: 23 April 2018
© The Author(s) 2018

Abstract
Hand, foot, and mouth disease (HFMD) is a global infectious disease resulting in millions of cases and even hundreds of deaths. Although a newly developed formalin-inactivated EV71 (FI-EV71) vaccine is effective against EV71, which is a major pathogen for HFMD, no vaccine against HFMD itself has yet been developed. Therefore, establishing a sensitive and accurate early warning system for HFMD is important. The early warning model for HFMD in the China Infectious Disease Automated-alert and Response System combines control chart and spatial statistics models to detect spatiotemporal abnormal aggregations of morbidity. However, that type of early warning for HFMD just involves retrospective analysis. In this study, we apply a Bayesian belief network (BBN) to estimate the increased risk of death and severe HFMD in the next month based on pathogen detection and environmental factors. Hunan province, one of the regions with the highest prevalence of HFMD in China, was selected as the study area. The results showed that compared with the traditional early warning model for HFMD, the proposed method can achieve a very high performance evaluation (the average AUC tests were more than 0.92). The model is also simple and easy to operate. Once the structure of the BBN is established, the increased risk of death and severe HFMD in the next month can be estimated based on any one node in the BBN.

Keywords Hand, foot, and mouth disease · Early warning · Spatiotemporal abnormal aggregation · Moving percentile method · Bayesian belief network

1 Introduction

Hand, foot, and mouth disease (HFMD) is a viral illness common in infants and children younger than 5 years old. Since the first HFMD case was reported in New Zealand in 1957, many parts of the world have reported epidemics of HFMD. The Asia Pacific Region is one of the major HFMD-endemic regions. Most countries there have reported HFMD epidemics in the past 30 years, including Australia (Burry et al. 1968), Japan (Daisuke and Masahiro 2011), South Korea (Ryu et al. 2010), Malaysia (Chua and Kasri 2011), Singapore (Ang et al. 2009), Thailand (Puenpa et al. 2011), Vietnam (Tu et al. 2007), and China

Yilan Liao and Bing Xu contributed equally to this work.

✉ Yilan Liao
liaoyl@lreis.ac.cn

✉ Lidong Gao
810173358@qq.com

[1] The State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

[2] The Key Laboratory of Surveillance and Early-Warning on Infectious Disease, Chinese Center for Disease Control and Prevention, Beijing 102206, China

[3] Sino-Danish College, University of Chinese Academy of Sciences, Beijing 100190, China

[4] School of Information Engineering, China University of Geosciences, Beijing 100083, China

[5] Hunan Provincial Center for Disease Control and Prevention, Changsha 410005, China

[6] Sino-Danish Center for Education and Research, Beijing 100190, China

(Xing et al. 2014a, b). Although HFMD is a mild viral infection, children under 5 years of age are particularly susceptible to the neurologic complications of this infection, which can result in encephalitis and death. During 2008–2013, 9,048,244 probable cases of HFMD were reported to the Chinese Center for Disease Control and Prevention (China CDC), of which 355,563 (3.9%) were laboratory confirmed, 92,503 (1.02%) were severe, and 2,707 (0.03%) were fatal (Xing et al. 2014a, b). Pathogen detection results have shown that EV71 causes approximately 50% of cases of HFMD in China and that it is a significant cause of morbidity and mortality (Pallansch and Roos 2007; Lin et al. 2002). Currently, there is no available vaccine against HFMD, even though a newly developed formalin-inactivated EV71 (FI-EV71) vaccine has been tested in clinical trials and has shown efficacy against EV71 (Tsou et al. 2015). Therefore, the early detection of the aberration of infectious disease occurrence and the estimated disease outbreak risk is significant for slowing down the epidemic spread and reducing the morbidity and death caused by the disease.

Some methods for forecasting disease outbreaks of HFMD have been successfully applied in some countries and regions. In Japan, an index calculated from the number of cases per week per sentinel medical institution was used as a disease warning threshold. If the index in an area exceeds the critical value for the onset of an epidemic and continues until the index in that area is lower than the critical value for the end of an epidemic, it can predict that there is an unusual aberration of HFMD cases (Murakami et al. 2004). In China, a nationwide web-based automated system for outbreak detection and rapid response was developed in 2008. In the China Infectious Disease Automated-alert and Response System (CIDARS), the fixed-threshold detection method, the temporal detection method, and the spatial detection method were applied to detect three disease aberrations (Yang et al. 2011). HFMD surveillance has been included in CIDARS. In some regions of China with a high HFMD prevalence, researchers have proposed many statistical models to estimate local development trends of HFMD. These research results give support for local early warnings of HFMD. In Shandong province, a conditional logistic regression was used to explore an early warning index of HFMD death based on clinical treatment, past medical history, clinical symptoms, and signs of ill children (Liu et al. 2013). A hybrid model combining a seasonal autoregressive integrated moving average (ARIMA) model and a nonlinear autoregressive neural network (NARNN) was proposed to predict the expected incidence of cases in Shenzhen city from December 2012 to May 2013 (Yu et al. 2014). In Dalian city, a negative binomial regression model was used to estimate the weekly baseline number of HFMD cases and identified the optimal alerting threshold between different tested threshold values during the epidemic and non-epidemic year (An et al. 2016). Most of these early warning models use historical case data to forecast an imminent outbreak; in fact, these models use only retrospective research (Edmond et al. 2011). The operation of these models also closely depends on the data integrity of the risk factor variables. However, in many cases, the risk factor variables are difficult to collect completely in time.

A Bayesian belief network (BBN) is a graphical model for probabilistic relationships among a set of variables; it can express the mutual dependencies between variables in terms of quality and quantity, and it provides a good explanation for knowledge representation and reasoning (Holický et al. 2013). This method can also produce uncertainty estimates even with missing values. Because of its simplicity and soundness, BBN has been used in many fields, especially in medical and industrial diagnosis (Liao et al. 2010; Liu et al. 2012). So this study establishes an early warning model based on BBN to predict the outbreak risk of severe HFMD and death. Hunan province, one of the regions with the highest prevalence of HFMD in China, was selected as the study area. According to previous study, children under 5 years old accounted for most of the HFMD cases (Deng et al. 2013), and EV71 predominated in laboratory-confirmed cases, accounting for 45% of mild, 80% of severe, and 93% of fatal cases (Pallansch and Roos 2007; Lin et al. 2002). In order to analysis the potential seasonal drivers of HFMD, social-economic factors such as gross domestic product and gross regional product were also put into account (Xing et al. 2014a, b). Meanwhile, a variety of studies identified that meteorological factors including temperature, rainfall, relative humidity, air pressure, and hours of sunshine all have a strong association of the prevalence of HFMD (Deng et al. 2013; Xing et al. 2014a, b; Yu et al. 2014). Therefore, in the proposed model, the monthly pathogen detection ratio, demographic factors, socioeconomic factors, and monthly meteorological factors are all taken into account to estimate the outbreak risk of death and severe HFMD in the next month. Based on the characteristic of BBN, once the structure is established and parameters are learned, the risk of death and sever HFMD can be estimated based on any one node (factor) in the BBN, which can handle the situation where the data of some variables are unavailable.

2 Data and method

2.1 Study area

In this study, Hunan, a province in south central China, was selected as the study area. Hunan is located south of the

middle course of the Yangtze River and south of Lake Dongting, situated between the 108°47′–114°16′ east longitudes and the 24°38′–30°08′ north latitudes. It covers an area of 211,800 square kilometers and is divided into 14 cities. Mountains and hills occupy more than 80% of the area, with plains comprising less than 20% of the whole province. Hunan's climate is subtropical, where sunshine and rainfall are both abundant concentrated. In spring, the temperature would fluctuate in a wide range, whereas in summer and autumn, it is pretty hot and humid, and in winter it's usually cool and damp. According to the meteorological data, the annual average temperature of Hunan is around 16–19 degree centigrade, with average 38 °C (37–46 °F) in January and average 27–30 °C (81–86 °F) in July. The average annual precipitation is 1200–1700 mm and has an uneven distribution (information from Wikipedia article, "Hunan") (Fig. 1).

In terms of the socioeconomic condition, according to the statistical yearbook of Hunan province, during the study period, there were around 66,380,000 people living in Hunan with a GDP of 32903.71 RMB per capita, and the people living in cities accounted for 46.65% of the total population. Since its geographical environment advantages and rich natural resources, the development of Hunan's economy is good.

During 2009–2013, 547,027 HFMD cases were reported in Hunan. The 5-year prevalence reached 167.83 per 100,000 persons. There were 5,638 severe cases (1.03% of all cases). There were 296 cases of death, a mortality rate of 0.05%. The annual HFMD prevalence varied significantly over these 5 years, ranging from 287.32 (in 2012) to 54.31 per 100,000 persons (in 2009). According to the results of previous studies, there was also a significant difference in HFMD prevalence among the 14 cities in Hunan. The HFMD prevalence in Hunan also showed a seasonal trend in 2009–2013; April to July and September to November are the peak months of HFMD in Hunan (Yang et al. 2015). Hence, a clustering method for panel data was used to classify experimental scenarios based on (1) the monthly rate, severity, and mortality of HFMD in the all cities during the study period; (2) the monthly increase in the rate, severity, and mortality in the all cities during the study period; and (3) the rate of change of the rate, severity, and mortality increment in the all cities during the study period. Four experimental scenarios included a risk assessment of death and severe HFMD in the next month (1) in high-incidence cities during a peak period, (2) in high-incidence cities during a non-peak period, (3) in low-incidence cities during a peak period, and (4) in low-incidence cities during a non-peak period (Fig. 2).

2.2 Data sources

The period of the study was from January 1, 2011, to December 31, 2013, and the study unit was a city. The study included the rate, severity, and mortality of HFMD and the rate of pathogen EV71 detection in mild cases. People older than 5 years of age were ignored in the study, because HFMD mainly infects children under the age of five. In addition, considering that prekindergarten children and kindergarten students have different routes of exposure to pathogen EV71, we separately calculated the rates of pathogen EV71 detection in mild cases for children 0–3 years old and children 3–5 years old. All HFMD cases and pathogen detection laboratory data were collected from the local CDC.

The environmental factors in this study were meteorological and socioeconomic. In accordance with previous studies (Yang et al. 2015; Gao et al. 2014), the meteorological data mainly included the average monthly air pressure, maximum monthly wind speed, average monthly wind speed, average monthly relative humidity, minimum monthly relative humidity, and average monthly temperature, monthly temperature difference, monthly rainfall, and monthly sunshine. The meteorological data of all cities were obtained from the China Meteorological Data Sharing Service System and were interpolated by the inverse distance weighted interpolation method, based on the data of the 756 stations nationwide. The socioeconomic factors included the density of the susceptible population (i.e., children under the age of five) and urbanization levels. These data were extracted from the *Hunan Statistical Yearbook*. ArcGIS 10.0 was used to store the geographic data and display the results of this study.

2.3 Methodology

The framework for establishing an early warning model based on BBN to predict the outbreak risk of severe HFMD and death in Hunan comprises three parts: (1) defining the risk level; (2) building the BBN construct; and (3) calculating the joint probability of the target variable in the BBN and assessing the outbreak risk. In this study, the BBN software tool used was BN PowerSoft (Cheng and Greiner 1999).

2.3.1 Classification of risk level

In this study, we introduced the dynamic control chart percentile method (Yang et al. 2014; Xiong et al. 2013) to determine the risk level of HFMD outbreak; this model had been successfully used in CIDARS. In the dynamic control chart percentile method, to account for the

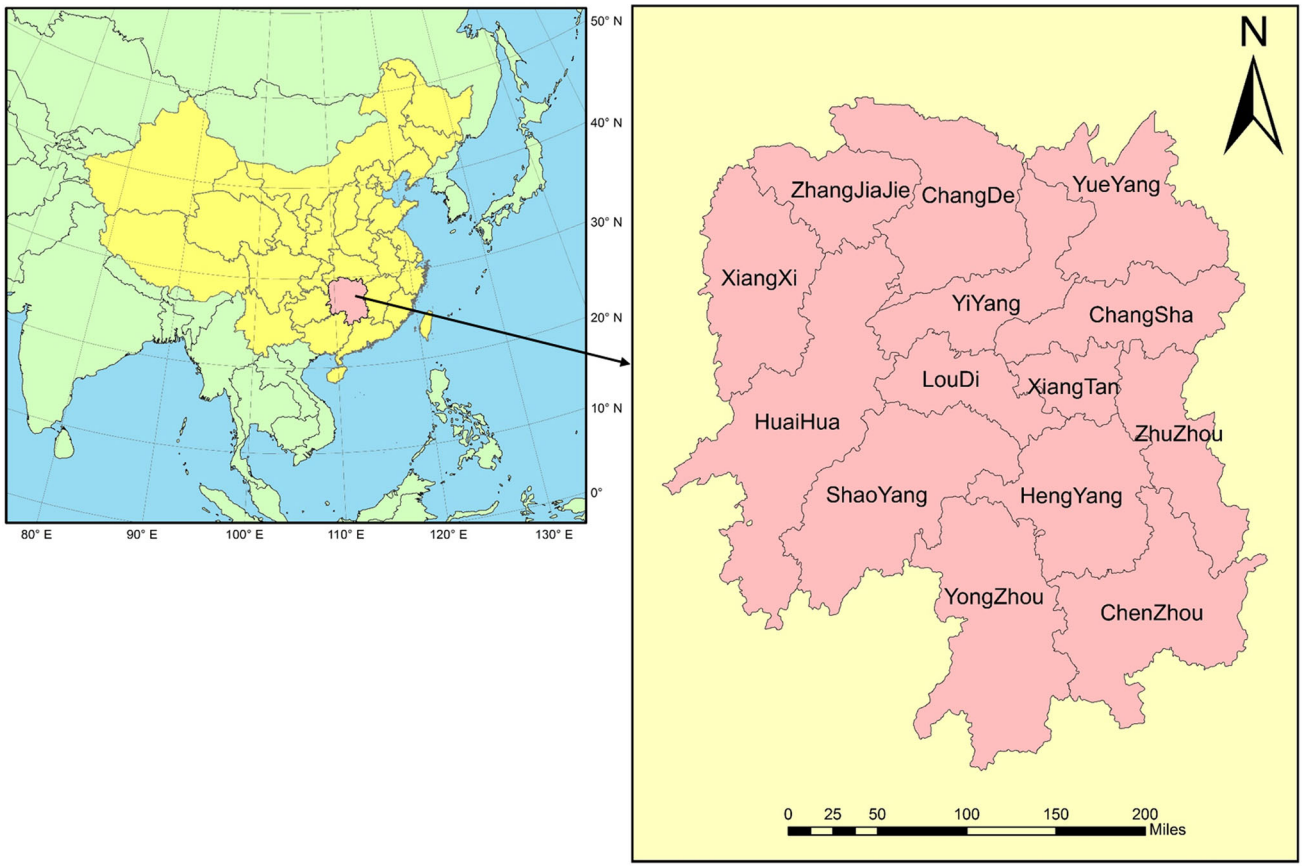


Fig. 1 Location of Hunan province in China

stability of the data, the most recent one-month period is used as the current observation period. The corresponding historical period included the same one-month period, the preceding one-month period, and the following 1-month period over the preceding 3 years of historical data. The current observation period and historical data block were dynamically moved forward month by month.

The reported death and severe incidence in the nine blocks of historical data were first arranged in ascending order, $X_1, X_2, \ldots, X_d, \ldots, X_n (n = 9)$. Then the percentile of the nine blocks of historical data was set as the indicator of potential aberration; p is a specified percentile of historical data, and X_P can be calculated as follows:

$$\mathrm{d} = 1 + (n-1)p \tag{1}$$

$$X_P = X([d]) + (X([d+1]) - X([d]))(d - [d]) \tag{2}$$

([] means take the integer of the number.)

Finally, the method defined the outbreak risk level of severe HFMD and death by comparing the reported death and severe incidence in the current 1-month period to different control limits of the corresponding historical period at the city level. That is, the risk was regarded as high and the risk level could be set at 2 if the actual incidence exceeded the upper control limit (P_{80}). If the actual incidence was greater than P_{50} and less than P_{80}, the risk was moderate, and the risk level was set as 1. If the actual incidence was less than P_{50}, the risk level was set at 0. The selection of these threshold values takes the ones used in CIDARS as the standard.

2.3.2 Building the BBN construct

A Bayesian network is represented by $\mathrm{BN} = \langle N, A, \Theta \rangle$, where $\langle \mathrm{N}, \mathrm{A} \rangle$is a directed acyclic graph (DAG)—each node $\mathrm{n} \in \mathrm{N}$ represents a domain variable (corresponding perhaps to a database attribute), and each arc $\mathrm{a} \in \mathrm{A}$ between nodes represents a probabilistic dependency between the associated nodes. Associated with each node $n_i \in N$ is a conditional probability distribution (CPtable), collectively represented by $\Theta = \{\theta_i\}$, which quantifies how much a node depends on its parents (see Pearl 1988). In this study, the BBN is presented composed of a set of nodes, representing the risk variables of severe HFMD and death, and a set of edges, representing the probabilistic causal dependence among the variables. Before inputting the variables into the BBN, continuous variables were discretized by the entropy method. In the entropy method, the minimum entropy value in a numerical attribute is chosen as a split

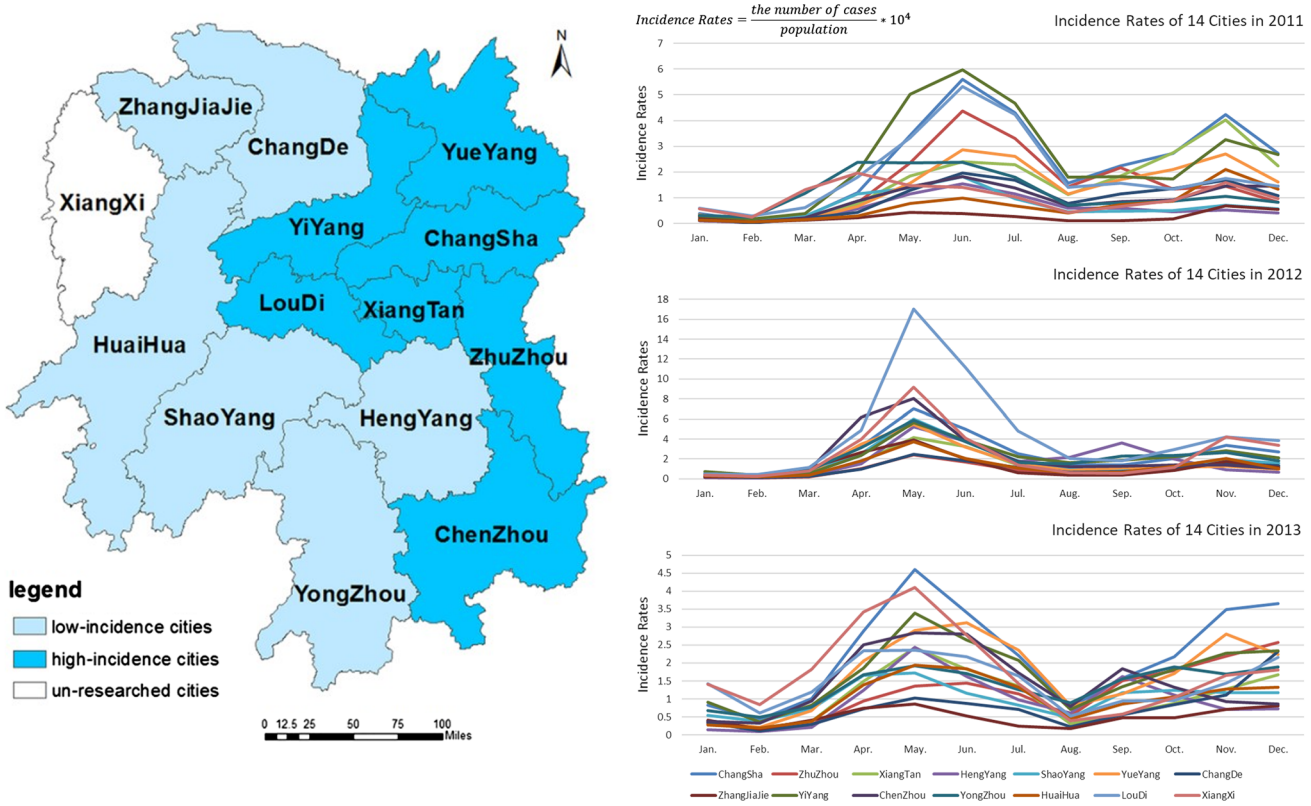


Fig. 2 High-incidence and low-incidence cities in Hunan (Xiangxi is not included due to low data quality)

point, and then the classified results of the intervals recursively and eventually gain discretized results.

In the BBN, the nodes with edges directing into them are called ''child'' nodes, and the nodes from which the edges depart are called ''parent'' nodes. That is, if there is an edge from node N1 to another node N2, N1 is called the parent of N2. Nodes without arches directing into them are called ''root'' nodes. Therefore, a proper BBN network construct is very significant for assessing a future increase in the outbreak risk of severe HFMD and death, based on the BBN.

Generally, three methods—the expert knowledge method, the network structure learning algorithm, and an integration of the above two methods—are applied to build the Bayesian belief network. The expert knowledge method builds a network according to experts' comments and the results of literature reviews, while the network structure learning algorithm is based on the independence test proposed by Cheng and Greiner (1999). The network structure learning algorithm has three phases: drafting, thickening, and thinning (Liao et al. 2017). In the first phase, the algorithm computes the mutual information of each pair of nodes as a measure of closeness, and it creates a draft based on this information. In the second phase, the algorithm adds arcs when the pairs of nodes cannot be d-separated. The node X is said to be d-separated from node Y if there is no active undirected path between X and Y. The result of the second phase is an independence map (I-map) of the underlying dependency model. In the third phase, if an open path separate from the current arc remains, it will be brought from the set of directed arcs joining vertices. Then the independence of the corresponding block set that blocks each open path between these two nodes by a set of minimum number of nodes is tested. The arc will be removed from the I-map if the two nodes of the arc can be d-separated. The result of the third phase is the minimal I-map. In the study, we applied an integration of the expert knowledge method and the network structure learning algorithm to construct the network.

2.3.3 Assessing the outbreak risk of severe HFMD and death

After the structure of Bayesian network was constructed (i.e., determining what depends on what), the next step is learning the parameters (i.e., the strength of these dependencies, as encoded by the entries in CPtables). For each BBN node, a conditional probability formula or table, which represents the probabilities of each value of the node given the conditions of its parents, is supplied. Meanwhile, the distributions of the parents can be calculated given the values of their children (Cooper and Herskovits 1992).

After the conditional probability table is obtained, the joint probability distributions can be decomposed into local probability distributions (Liu et al. 2012). Therefore, in the study, the assessment of the outbreak risk of severe HFMD and death was performed simply by calculating the conditional probabilities among the variables, based on their cause and consequence relationships. The conditional probability distribution in the BBN can be expressed as follows:

$$\begin{aligned} \mathrm{P}(X_1 = x_1, \ldots, X_n = x_n) &= \mathrm{P}(x_1, \ldots, x_n) \\ &= \prod_{i=1}^{n} P(x_i | Parents(X_i)) \end{aligned} \quad (3)$$

where $\mathrm{P}(x_1, \ldots, x_n)$ refers to the probability of a specific combination of values $x_1, \ldots, x_n$ from the set of variables $X_1, \ldots, X_n$, where Parent (X_i) refers to the set of $X_i's$ immediate parent nodes. Thus, $P(x_i|Parents(X_i))$ reflects the conditional probability, which is related to the node X_i based on its parent nodes (Liao et al. 2010).

3 Results

3.1 BBN structure for assessing the outbreak risk of severe HFMD and death

As described in Sect. 2.1, four BBN structures were applied to assess the risk of death and severe HFMD in the next month in different experimental scenarios. Before entering the data into the models, all the data used in this research should be discretized. Table 1 shows the results of each variable classification attribute in the BBN. The experimental scenarios included (1) in high-incidence cities during a peak period, (2) in high-incidence cities during a non-peak period, (3) in low-incidence cities during a peak period, and (4) in low-incidence cities during a non-peak period. In each experimental scenario, 70% of the surveillance data was used to build the early warning model for the outbreak of severe HFMD and death in the next month. Thirty percent of the surveillance data was applied to evaluate the accuracy of the model.

Figure 3 illustrates the network structures of the BBN used to estimate the outbreak risk of severe HFMD and death in the next month in the four experimental scenarios. In each network structure, we selected the variable ''Risklevel'' as the parent node of the various factors. All independent variables were represented by nodes; the arcs between the nodes indicated the relationships of the variables. As seen in Fig. 3, the effect on the outbreak risk of severe HFMD and death in the next month varied among the selected factors. However, in all the experimental scenarios, there were some common risk factors for this outbreak risk. The common risk factors included the total current monthly rates of pathogen EV71 detection in mild cases (the variable "EV71_rate"), the current monthly rates of pathogen EV71 detection for children aged 0–3 years (the variable "EV71_rate_1") and aged 3–5 years (also the variable "EV71_rate_1"). In addition to the influence of the spread of pathogen EV71, some meteorological variables, including the current maximum monthly wind speed (the variable "MaxWinSpeed"), the average monthly temperature (the variable "AveTemp"), and the average monthly relative humidity (the variable "AveReHumidy"), affected the outbreak risk of severe HFMD and death in the next month in the high-incidence cities. During the peak period, the outbreak risk of severe HFMD and death in the next month in all cities was also related to the current monthly temperature and the maximum monthly wind speed. Nevertheless, the meteorological factors that had a direct impact on the outbreak risk of severe HFMD and death in the next month during the non-peak time were changed into the current monthly rainfall (the variable "Rainfall") and the average monthly relative humidity. The influence of all socioeconomic factors on the outbreak risk of severe HFMD and death in the next month also varied in the different experimental scenarios.

In our BBN structures, the total current monthly rates of pathogen EV71 detection in mild cases was one of risk factors with the most direct influence on the outbreak risk of severe HFMD and death in the next month. Different from how they influenced this outbreak risk, socioeconomic factors mainly affected the total current monthly rates of pathogen EV71 detection in mild cases in all the experimental scenarios. However, meteorological factors also had an impact on the spread of pathogen EV71 in some experimental scenarios. For example, the total current monthly rates of pathogen EV71 detection in mild cases in the high-incidence cities were directly related to the current monthly rainfall. During the peak period, the current maximum monthly wind speed also influenced the total current monthly rates of pathogen EV71 detection in mild cases.

3.2 Outbreak risk assessment of severe HFMD and death in Hunan

As we constructed four BBN network structures for these four study areas, a probabilistic inference reflecting the dependence relationships between the risk level of severe HFMD and death and the relevant risk factors was essential. Table 2 pinpoints the conditional probability distribution of the four BBNs for estimating the HFMD outbreak risk in the four study areas. It is obvious that there are two dominant aspects.

Table 1 The discretization results of each classification attribute in the BBN

Area	Class	Rate of virus EV71 detection in mild cases at different ages (1/100)			The density of the susceptible population (pop/km2)	Urbanization level (%)	Average wind speed (m/s)	Maximum wind speed (m/s)
		0–5 years	0–3 years	3–5 years				
High-incidence cities during peak period	1	< 0.057	< 0.050	< 0.05	< 20.299	< 42.815	< 1.658	< 3.934
	2	0.057–0.151	0.050–0.155	0.05–0.354	20.299–21.081	42.815–48.555	1.658–1.797	3.934–4.222
	3	0.151–0.267	0.155–0.263	0.354–0.9	21.081–29.716	48.555–56.825	1.797–1.958	4.222–4.499
	4	0.267–0.407	0.263–0.415	> 0.9	29.716–33.449	56.825–69.78	1.958–2.194	4.499–4.943
	5	> 0.407	> 0.415	–	> 33.449	> 69.78	> 2.194	> 4.943
High-incidence cities during non-peak period	1	< 0.036	< 0.038	< 0.063	< 20.240	< 42.815	< 1.6029	< 3.498
	2	0.036–0.105	0.038–0.099	0.063–0.408	20.240–20.815	42.815–48.555	1.602–1.752	3.498–3.811
	3	0.105–0.223	0.099–0.219	> 0.408	20.815–29.183	48.555–56.825	1.752–1.946	3.811–4.150
	4	0.223–0.398	0.219–0.390	–	29.183–33.212	56.825–69.78	1.946–2.258	4.150–5.162
	5	> 0.398	> 0.390	–	> 33.212	> 69.78	> 2.258	> 5.162
Low-incidence cities during peak period	1	< 0.055	< 0.052	< 0.045	< 10.891	< 37.815	< 1.378	< 3.669
	2	0.055–0.103	0.052–0.090	0.045–0.348	10.891–14.482	37.815–39.985	1.378–1.624	3.669–4.175
	3	0.103–0.209	0.090–0.211	> 0.348	14.482–23.130	39.985–42.96	1.624–1.926	4.175–4.597
	4	0.209–0.398	0.211–0.387	–	23.130–32.872	42.96–47.445	1.926–2.325	4.597–5.196
	5	> 0.398	> 0.387	–	> 32.872	> 47.445	> 2.325	> 5.196
Low-incidence cities during non-peak period	1	< 0.031	< 0.034	< 0.042	< 10.891	< 37.815	< 1.339	< 3.229
	2	0.031–0.082	0.034–0.850	0.042–0.367	10.89–14.482	37.815–39.985	1.339–1.558	3.229–3.809
	3	0.082–0.148	0.850–0.156	> 0.367	14.482–23.130	39.985–42.96	1.558–1.951	3.809–4.298
	4	0.148–0.255	0.156–0.304	–	23.130–32.688	42.96–47.445	1.951–2.226	4.298–5.003
	5	> 0.255	> 0.304	–	> 32.688	> 47.445	> 2.226	> 5.003

Area	Class	Minimum relative humidity (%)	Average relative humidity (%)	Temperature diffrence (°C)	Average temperature (°C)	Average air pressure (hPa)	Sunshine hours (h)	Rainfall (mm)
High-incidence cities during peak period	1	< 49.535	< 70.850	< 7.213	< 18.283	< 982.689	< 21.916	< 3.017
	2	49.535–53.342	70.850–73.994	7.213–7.547	18.283–21.508	982.689–987.763	21.916–30.908	3.017–3.752
	3	53.342–55.675	73.994–75.727	7.547–8.029	21.508–23.229	987.763–994.911	30.908–41.319	3.752–5.751
	4	55.675–58.935	75.727–77.765	8.029–8.544	23.229–27.321	994.911–1003.306	41.319–52.896	5.751–7.574
	5	> 58.935	> 77.765	> 8.544	> 27.321	> 1003.306	> 52.896	> 7.574
High-incidence cities during non-peak period	1	< 47.044	< 68.548	< 5.501	< 5.209	< 992.643	< 11.114	< 1.214
	2	47.044–53.535	68.548–75.051	5.501–6.626	5.209–7.903	992.643–995.585	11.114–20.855	1.214–2.474
	3	53.535–59.416	75.051–78.433	6.626–7.795	7.903–12.745	995.585–1002.945	20.855–35.150	2.474–3.157
	4	59.416–64.612	78.433–80.549	7.795–8.743	12.745–29.159	1002.945–1014.528	35.150–70.844	3.157–5.044
	5	> 64.612	> 80.549	> 8.743	> 29.159	> 1014.528	> 70.844	> 5.044

Table 1 (continued)

Area	Class	Minimum relative humidity (%)	Average relative humidity (%)	Temperature diffrence (°C)	Average temperature (°C)	Average air pressure (hPa)	Sunshine hours (h)	Rainfall (mm)
Low-incidence cities during peak period	1	< 49.743	< 71.699	< 7.224	< 17.818	< 972.202	< 27.755	< 2.973
	2	49.743–52.175	71.699–74.191	7.224– < 7.753	17.818–– 21.035	972.202–978.225	27.755–40.876	2.973–3.998
	3	52.175–55.789	74.191–76.113	7.753–8.325	21.035–22.971	978.225–984.156	40.876–51.502	3.998–5.778
	4	55.789–58.959	76.113–78.677	8.325–9.143	22.971–26.720	984.156–992.259	51.502–69.303	5.778–7.126
	5	> 58.959	> 78.677	> 9.143	> 26.720	> 992.259	> 69.303	> 7.126
Low-incidence cities during non-peak period	1	< 46.400	< 68.861	< 5.281	< 5.083	< 977.466	< 10.691	< 0.936
	2	46.400–52.067	68.861–73.886	5.28–6.289	5.083–8.314	977.466–987.020	10.691–23.567	0.936–1.714
	3	52.067–57.340	73.886–77.061	6.289–7.823	8.314–12.314	987.020–989.873	23.567–34.510	1.714–2.628
	4	57.340–61.711	77.061–79.274	7.823–9.136	12.314–28.404	989.873–1000.694	34.510–81.942	2.628–4.348
	5	> 61.711	> 79.274	> 9.136	> 28.404	> 1000.694	> 81.942	> 4.348

First, the primary variables that impact the risk levels differ to a great extent in these four different areas, and various probabilities are inferred by the Bayesian posterior probability estimation. In other words, there are various probabilities of the different values of the variables in estimating the risk levels (i.e., the three risk levels 0, 1, 2). In high-incidence cities during the peak period, as in the BBN structure described above, the variables that most notably produce an effect on the risk level of severe HFMD and death are the rate of pathogen EV71 detection in mild cases for children 0–5 and 3–5 years old, the urbanization level, the average wind speed, the maximum wind speed, the average relative humidity, the temperature difference, the average temperature, the average air pressure, and sunshine hours. The probabilistic inference results of all these relative variables linked in the above structure vary because of their different influences on the HFMD risk level. For instance, when the value of the rate of pathogen EV71 detection in mild cases for children aged 3–5 is in class one (< 0.05), the risk levels for 0, 1, and 2 are 0.422, 0.305, and 0.388, respectively, so we can conclude that the probability of no risk in this area is greater. However, when the values of the rate of pathogen EV71 detection for children 3–5 years old increase, the probabilistic inference results of all risk level are reduced, which means that the higher the rate of pathogen EV71 detection for children 3–5 years old, the lower the risk forecast for any risk level.

A similar situation occurs with other influencing variables, such as urbanization level, while the average temperature variable is just the opposite. When the value of this variable is in class 5 (> 27.321), the probability experiences a peak at the risk level of 0, with 0.462, which means that the areas with these temperatures are safe. In high-incidence cities during a non-peak period, the variables that most notably produce an effect on the risk level of severe HFMD and death are the rate of pathogen EV71 detection in mild cases for children aged 0–3 and 3–5, the urbanization level, the average relative humidity, and rainfall. Similarly, the probabilistic inference results of these relative variables may vary. If the rainfall is in class 4 (3.157–5.044), the probability will be 0.545 at a risk level of 2, so these places should be aware of the corresponding protective measures taken by CDC or governments. In low-incidence cities during a peak period, the variables are the rate of pathogen EV71 detection in mild cases for children 0–5 and 3–5 years old, the average temperature, and the average air pressure. In low-incidence cities during a non-peak period, the variables are the rate of pathogen EV71 detection in mild cases for children aged 0–5, the rate of pathogen EV71 detection in mild cases for children aged 3–5 years, the urbanization level, the average wind speed, the average relative humidity, the temperature difference, the average air pressure, and rainfall.

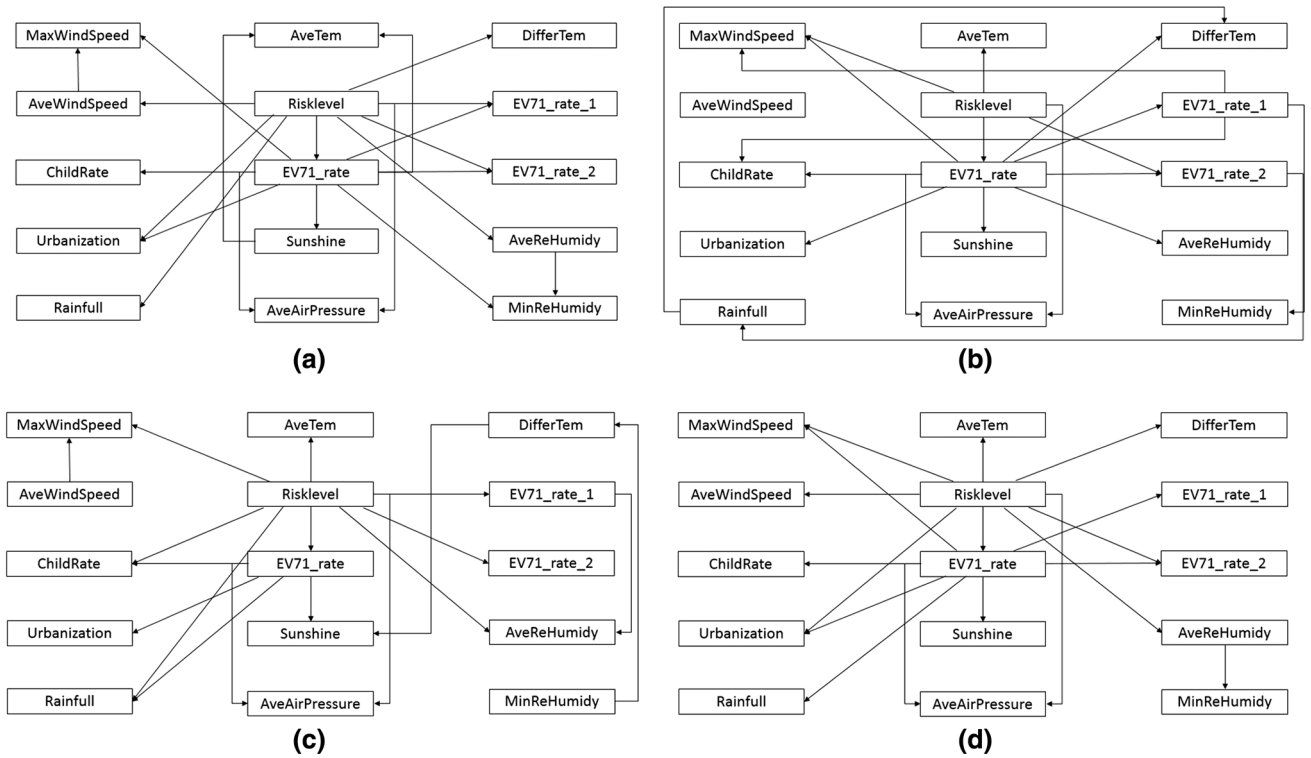


Fig. 3 BBN structures **a** in high-incidence cities during a peak period, **b** in high-incidence cities during a non-peak period, **c** in low-incidence cities during a peak period, and **d** in low-incidence cities during a non-peak period

Secondly, the same variable can have a different impact on different areas. In Table 2, we can see that the rate of pathogen EV71 detection in mild cases for children aged 0–5 has an impact in all four study areas, while other variables impact just some areas. For example, rainfall has effects on both high-incidence and low-incidence cities during a non-peak period. In addition, in all areas, the density of the susceptible population and the minimum relative humidity have no relationship with the outbreak risk assessment of severe HFMD and death, so they have been ignored in Table 2. Otherwise, the probabilities of the same variable will vary in different areas.

Take a factor, Average Relative Humidity (ARH), as an example: For high-incidence cities during peak period area, when the value of ARH falls in "70.850–73.994", it belongs to the "class 2" according to the discretization results in Table 1. Then, we can utilize Table 2, which shows that the probability of risk level 0 is 0.333, the probability of risk level 1 is 0.176, and the probability of risk level 2 is 0.175, to predict the most probable risk level of HFMD outbreak in high-incidence cities during peak is 0. Meanwhile, for high-incidence cities during non-peak period area, when the ARH value is in class 2 (47.044–53.535), the probability of risk level 2, 0.455, is the highest. Other variables experience the same situation.

3.3 Accuracy analysis

The model performance was evaluated with receiver operating characteristic (ROC) analysis, which has been widely used in evaluating the predictive accuracy of a binary classifier. A typical ROC curve plots the true positive rate (sensitivity) against the false positive rate (1-specificity) for the entire range of possible thresholds, thus providing a unified representation for assessing the overall model performance. The area under the curve (AUC) can be used as a single performance measure to decide whether the model prediction is better than random (0.5). A perfect model will yield an AUC value of 1. Training AUC values is fairly high across models and is higher than the testing AUC values, as anticipated. The testing AUC values, which demonstrate the model's actual predictive powers, suggest whether the model predictions are better than random.

Because the ROC curve applies only to binary classifiers, we need to reclassify the HFMD outbreak risk level (0 or 1 or 2) into two classes. Situation 1: we define level = 0 as the class that means there is no HFMD outbreak, and we merge level = 1 and level = 2 as the other class to represent the outbreak. Situation 2: we merge level = 0 and level = 1 to signify that the epidemic is not

Table 2 Conditional probability distribution table of the BBN for estimating HFMD outbreak risk in Hunan different areas (Density of the susceptible population and Minimum Relative Humidity has no relationship with the outbreak risk assessment of HFMD severe and death in all areas, so they have been ignored in this table.)

Area	Risk level	Class	Probability of different variables										
			Rate of virus EV71 detection in mild cases at different ages (1/100)			Urbanization level (%)	Average wind speed (m/s)	Maximum wind speed (m/s)	Average relative humidity (%)	Temperature difference (°C)	Average temperature (°C)	Average air pressure (hPa)	Sunshine hours(h)	Rainfall(mm)
			0–5 years	0–3 years	3–5 years									
High-incidence cities during peak period	0	1	0.256	–	0.422	0.204	0.231	0.135	0.128	0.128	0.026	0.284	0.105	–
		2	0.179	–	0.261	0.262	0.128	0.207	0.333	0.282	0.051	0.137	0.203	–
		3	0.256	–	0.153	0.251	0.128	0.221	0.308	0.282	0.205	0.189	0.126	–
		4	0.103	–	0.164	0.166	0.282	0.215	0.179	0.231	0.256	0.205	0.172	–
		5	0.205	–	–	0.118	0.231	0.223	0.051	0.077	0.462	0.185	0.394	–
	1	1	0.118	–	0.305	0.223	0.235	0.224	0.176	0.235	0.118	0.279	0.171	–
		2	0.294	–	0.248	0.194	0.235	0.208	0.176	0.059	0.059	0.162	0.224	–
		3	0.176	–	0.249	0.206	0.235	0.219	0.059	0.294	0.353	0.187	0.189	–
		4	0.235	–	0.198	0.219	0.176	0.184	0.294	0.294	0.412	0.157	0.226	–
		5	0.176	–	–	0.157	0.118	0.165	0.294	0.118	0.059	0.215	0.191	–
	2	1	0.175	–	0.388	0.249	0.158	0.218	0.246	0.281	0.298	0.128	0.294	–
		2	0.246	–	0.297	0.180	0.263	0.186	0.175	0.140	0.333	0.225	0.211	–
		3	0.263	–	0.234	0.221	0.298	0.210	0.193	0.123	0.193	0.271	0.200	–
		4	0.140	–	0.080	0.260	0.123	0.244	0.140	0.175	0.140	0.193	0.223	–
		5	0.175	–	–	0.089	0.158	0.142	0.246	0.281	0.035	0.184	0.072	–
High-incidence cities during non-peak period	0	1	0.245	0.245	–	0.245	–	–	0.340	–	–	–	–	0.302
		2	0.208	0.208	–	0.245	–	–	0.151	–	–	–	–	0.283
		3	0.226	0.170	–	0.208	–	–	0.189	–	–	–	–	0.132
		4	0.151	0.208	–	0.245	–	–	0.226	–	–	–	–	0.170
		5	0.170	0.170	–	0.057	–	–	0.094	–	–	–	–	0.113
	1	1	0.200	0.200	–	0.300	–	–	0.100	–	–	–	–	0.100
		2	0.100	0.200	–	0.300	–	–	0.100	–	–	–	–	0.200
		3	0.200	0.300	–	0.200	–	–	0.300	–	–	–	–	0.400
		4	0.300	0.200	–	0.100	–	–	0.300	–	–	–	–	0.100
		5	0.200	0.100	–	0.100	–	–	0.200	–	–	–	–	0.200
	2	1	0.273	0.273	–	0.273	–	–	0.091	–	–	–	–	0.091
		2	0.273	0.182	–	0.182	–	–	0.455	–	–	–	–	0.182
		3	0.182	0.091	–	0.273	–	–	0.273	–	–	–	–	0.091
		4	0.091	0.182	–	0.182	–	–	0.091	–	–	–	–	0.545
		5	0.182	0.273	–	0.091	–	–	0.091	–	–	–	–	0.091

Table 2 (continued)

Area	Risk level	Class	Probability of different variables											
			Rate of virus EV71 detection in mild cases at different ages (1/100)			Urbanization level (%)	Average wind speed (m/s)	Maximum wind speed (m/s)	Average relative humidity (%)	Temperature difference (°C)	Average temperature (°C)	Average air pressure (hPa)	Sunshine hours(h)	Rainfall(mm)
			0–5 years	0–3 years	3–5 years									
Low-incidence cities during peak period	0	1	0.313	–	0.453	–	–	0.190	–	–	0.063	0.225	–	–
		2	0.219	–	0.259	–	–	0.233	–	–	0.094	0.299	–	–
		3	0.125	–	0.288	–	–	0.152	–	–	0.219	0.167	–	–
		4	0.156	–	–	–	–	0.239	–	–	0.281	0.170	–	–
		5	0.188	–	–	–	–	0.186	–	–	0.344	0.139	–	–
	1	1	0.143	–	0.410	–	–	0.274	–	–	0.048	0.237	–	–
		2	0.143	–	0.309	–	–	0.175	–	–	0.143	0.250	–	–
		3	0.238	–	0.281	–	–	0.175	–	–	0.381	0.196	–	–
		4	0.286	–	–	–	–	0.192	–	–	0.333	0.169	–	–
		5	0.190	–	–	–	–	0.184	–	–	0.095	0.148	–	–
	2	1	0.239	–	0.496	–	–	0.217	–	–	0.370	0.166	–	–
		2	0.196	–	0.245	–	–	0.167	–	–	0.283	0.162	–	–
		3	0.217	–	0.259	–	–	0.207	–	–	0.239	0.165	–	–
		4	0.109	–	–	–	–	0.193	–	–	0.087	0.244	–	–
		5	0.239	–	–	–	–	0.216	–	–	0.022	0.262	–	–
Low-incidence cities during non-peak period	0	1	0.238	–	0.524	0.210	0.167	–	0.214	0.214	–	0.204	–	0.286
		2	0.167	–	0.181	0.195	0.262	–	0.214	0.238	–	0.217	–	0.286
		3	0.286	–	0.295	0.238	0.238	–	0.214	0.262	–	0.179	–	0.167
		4	0.143	–	–	0.211	0.190	–	0.214	0.143	–	0.160	–	0.119
		5	0.167	–	–	0.147	0.143	–	0.143	0.143	–	0.239	–	0.143
	1	1	0.100	–	0.338	0.230	0.300	–	0.100	0.400	–	0.163	–	0.100
		2	0.200	–	0.314	0.206	0.200	–	0.100	0.200	–	0.163	–	0.100
		3	0.200	–	0.348	0.196	0.100	–	0.200	0.100	–	0.239	–	0.300
		4	0.200	–	–	0.163	0.200	–	0.300	0.200	–	0.230	–	0.400
		5	0.300	–	–	0.206	0.200	–	0.300	0.100	–	0.206	–	0.100
	2	1	0.231	–	0.443	0.180	0.385	–	0.077	0.154	–	0.147	–	0.154
		2	0.231	–	0.245	0.180	0.154	–	0.308	0.231	–	0.213	–	0.077
		3	0.077	–	0.311	0.246	0.154	–	0.154	0.154	–	0.180	–	0.231
		4	0.231	–	–	0.246	0.231	–	0.308	0.231	–	0.279	–	0.385
		5	0.231	–	–	0.147	0.077	–	0.154	0.231	–	0.180	–	0.154

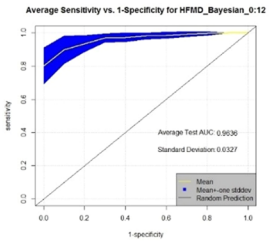
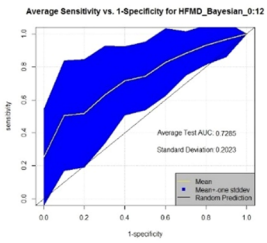
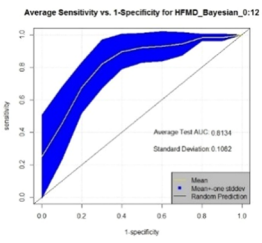
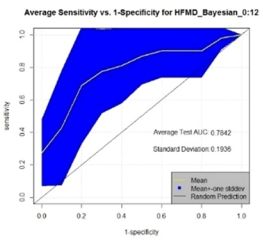
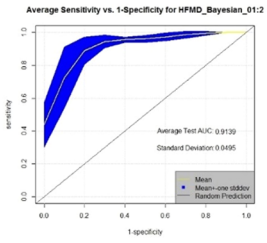
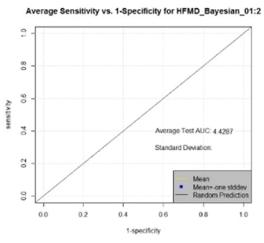
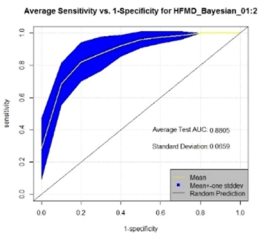
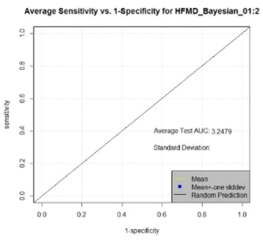
Reclassification	High-incidence cities during peak period	High-incidence cities during non-peak period	Low-incidence cities during peak period	Low-incidence cities during non-peak period
Situation 1: Level 0 as class 0; level 1 and 2 as class 1.	 AUC:0.9636 SD:0.0327	 AUC:0.7285 SD:0.2023	 AUC:0.8134 SD:0.1082	 AUC:0.7842 SD:0.1936
Situation 2: Level 0 and 1 as class 0; level 2 as class 2.	 AUC:0.9139 SD:0.0495	 AUC: SD:	 AUC:0.8805 SD:0.0659	 AUC: SD:

Fig. 4 ROC curve for belief network models applied in four experimental scenarios

serious, and level = 2 is a representation that the HFMD outbreak risk is very high.

As can be seen from Fig. 4, which clearly illustrates the performance of the BBN models applied in the four experimental scenarios, in terms of 10 times the average testing AUC value, no matter how the risk levels were reclassified, the models achieved the best predictive accuracy in the first scenario (high-incidence cities during a peak period), where the average testing AUC value was 0.9636 for reclassification 1 and 0.9139 for reclassification 2. This was followed by the third scenario (low-incidence cities during a peak period) and the fourth scenario (low-incidence cities during a non-peak period). The second scenario (high-incidence cities during a non-peak period) produced a comparatively low AUC value of 0.7285, which was also actually a good performance value. In terms of the standard deviation, which illustrates the stability of the predictive precision, the same order can be found, in which the most stable predictive accuracy applied to the first scenario, followed by the third and fourth scenarios; the second scenario produced the highest standard deviation value, and its predictive accuracy was the most unstable.

A conclusion that can be drawn from the analysis above is that whether the situation involves a peak or non-peak period has a crucial influence on the performance of the BBN Model. In a peak period, the model's AUC value is high and the SD value is low, which means an outstanding predictive performance. Otherwise, in a non-peak period, the model achieves a comparatively low AUC value and a high SD value, which means that the performance of the model is not as good, and the predictive result is very unstable.

Meanwhile, as a comparison with other predictive models, we also applied the logistic regression model, which is a strong scientific method utilizing one or more variables to forecast another dependent variable value. We also applied the rough set method, proposed by Pawlak in 1982, which has been used successfully in diagnosis and outcome prediction, for the same data of high-incidence cities during a peak period. In order to verify the accuracy and reliability of the different models, we process bootstrap resampling for 30 times, producing 30 sets of experimental data, and input them to each model and evaluated the predictive results through the ROC curve. The ROC curves

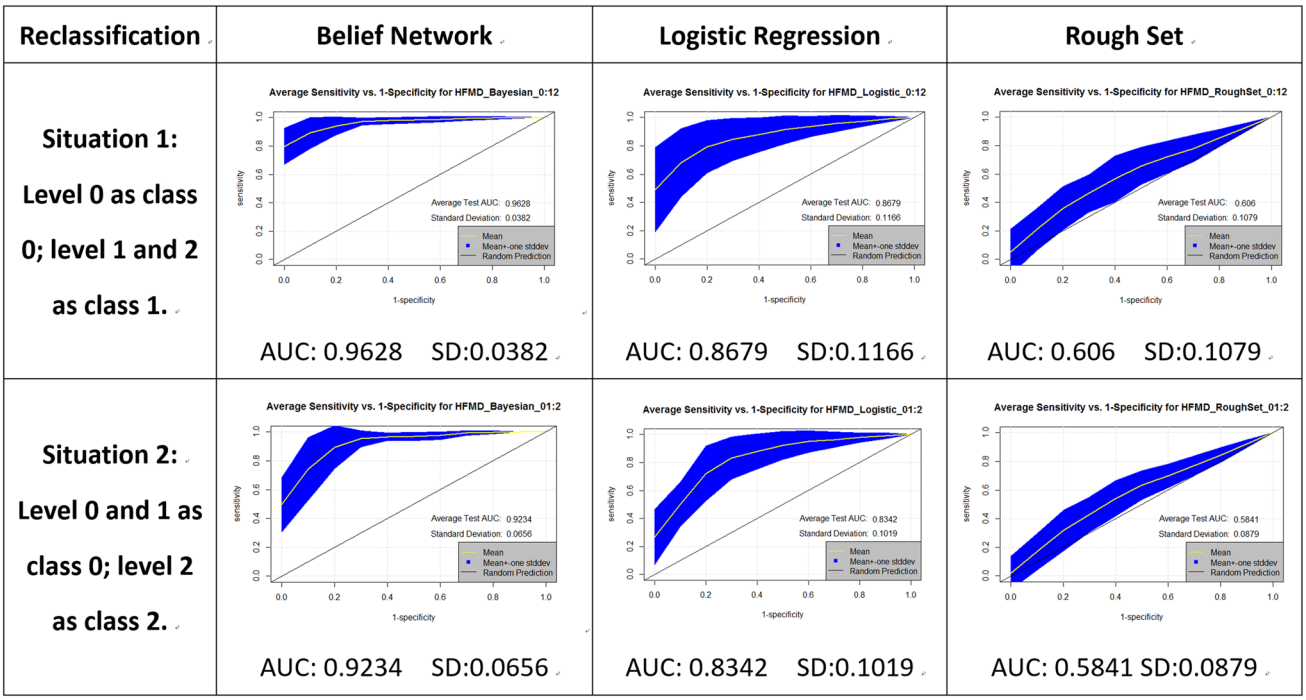


Fig. 5 ROC curve for Bayesian belief network model, logistic regression model, and rough set method prediction averaged on each of 30 replicate runs. (The 1:1 line indicates the condition if the prediction is completely out of random chance. (AUC = 0.5))

for each model's average prediction of 30 replicate runs, with a mean test AUC value and corresponding standard deviation, are provided in Fig. 5.

In the lateral view, no matter how the risk level was reclassified, the Bayesian network model achieved a very high performance evaluation (the average AUC test was 0.9628 for situation 1 and was 0.9234 for situation 2) and was better than the logistic regression model (where the average AUC test was 0.8679 for situation 1 and 0.8342 for situation 2) and the rough set method (where the average AUC test was 0.606 for situation 1 and 0.5841 for situation 2). In the longitudinal view, the stability of the Bayesian network was good (the standard deviation was 0.0382 for situation 1 and was 0.0656 for situation 2), regardless of the classification of the disease, and the logistic regression model (where the standard deviation was 0.1166 for situation 1 and was 0.1019 for situation 2) and the rough set method (where the standard deviation was 0.1079 for situation 1 and was 0.0879 for situation 2) did not have this advantage.

4 Discussion and conclusions

In this study, an early warning model was developed to predict the outbreak risk of severe HFMD and death in Hunan Province, China. In accordance with the temporal spatial epidemic trend of HFMD in Hunan between January 2010 and December 2013, the study was divided into four experimental scenarios. In each scenario, the proposed model applied probabilistic inference based on a BBN structure to assess the outbreak risk of severe HFMD and death in the next month. In four BBN structures, the relationships among outbreak risk, the pathogen detection rate, demographics, and socioeconomic and meteorological factors varied. ROC analysis was applied to evaluate the model's performance. The results showed that the performance of the proposed method was better than those of the rough set algorithm and the logistic regression.

Different from previous early warning models for HFMD, the proposed model used a network to express the mutual dependencies between the outbreak risk of severe HFMD and death and the risk factors. This clear representation of variable relationships will help us to understand and explore the causes of outbreaks of severe HFMD and death. More importantly, this model can make uncertainty estimates even with missing values. Although the proposed method was tested only in Hunan, it can be used as an early warning tool to solve HFMD outbreak prevention and control problems in any region, since BBN can provide a simple and unified method to address different problems in different areas.

However, there are also some limitations. The BBN in the proposed model can only deal with categorical data, so all continuous data needs to be discretized before being input to the model. This data discretization may lead to information loss. Moreover, while an accurate network is significant to the accuracy of the model, it is difficult to choose suitable network construction algorithms. Because the transmission path of HFMD in different areas varied, it

was also hard to include all the risk factors in the study. Therefore, in a later study, we should take more risk factors of HFMD outbreaks into account.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Nos. 41471377, 41531179, 41421001), the Featured Institute Construction Services Program of the Institute of Geographic Sciences and Natural Resources Research, CAS (No. TSYJS03), the Project Coupling big data and sampling for spatial mapping (No. 088RA200YA) supported by LREIS, and the Scientific research project (No. 20101801) supported by the Chinese Preventive Medicine Association.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

An Q, Wu J, Fan X, Pan L, Wei S (2016) Using a negative binomial regression model for early warning at the start of a hand foot mouth disease epidemic in Dalian, Liaoning province, China. PLoS ONE 11(6):e0157815

Ang LW, Koh BK, Chan KP, Chua LT, James L, Goh KT (2009) Epidemiology and control of hand, foot and mouth disease in Singapore, 2001–2007. Ann Acad Med Singap 38:106–112

Burry JN, Moore B, Mattner C (1968) Hand, foot and mouth disease in South Australia. Med J Aust 2(18):812

Cheng J, Greiner R (1999) Comparing Bayesian network classifiers. In: Proceedings of the fifteenth conference on uncertainty in artificial intelligence. Morgan Kaufmann Publishers, pp 101–108

Chua KB, Kasri AR (2011) Hand foot and mouth disease due to enterovirus 71 in Malaysia. Virol Sin 26:221–228

Cooper GF, Herskovits E (1992) A Bayesian method for the induction of probabilistic networks from data. Mach Learn 9:309–347

Daisuke O, Masahiro H (2011) The influence of temperature and humidity on the incidence of hand, foot, and mouth disease in Japan. Sci Total Environ 410–411(411):119–125

Deng T, Huang Y, Yu S et al (2013) Spatial-temporal clusters and risk factors of hand, foot, and mouth disease at the district level in Guangdong province, China. PLoS ONE 8(2):e56943. https://doi.org/10.1371/journal.pone.0056943

Edmond M, Wong C, Chuang SK (2011) Evaluation of sentinel surveillance system for monitoring hand, foot and mouth disease in Hong Kong. Public Health 125(11):777–783

Gao LD, Hu SX, Zhang H, Luo KW, Liu YZ, Xu QH, Huang W, Deng ZH, Zhou SF, Liu FQ, Zhang F, Chen Y (2014) Correlation analysis of EV71 detection and case severity in hand, foot, and mouth disease in the Hunan province of China. PLoS ONE 9(6):e100003. https://doi.org/10.1371/journal.pone.0100003

Holický M, Marková J, Sýkora M (2013) Forensic assessment of a bridge downfall using Bayesian networks. Eng Fail Anal 30:1–9

Liao Y, Wang J, Guo Y et al (2010) Risk assessment of human neural tube defects using a Bayesian belief network. Stoch Environ Res Risk Assess 24:93. https://doi.org/10.1007/s00477-009-0303-5

Liao Y et al (2017) A new method for assessing the risk of infectious disease outbreak. Sci Rep 7:40084. https://doi.org/10.1038/srep40084

Lin TY, Chu C, Chiu CH (2002) Lactoferrin inhibits enterovirus 71 infection of human embryonal rhabdomyosarcoma cells in vitro. J Infect Dis 186:1161–1164

Liu KFR, Lu CF, Chen CW et al (2012) Applying Bayesian belief networks to health risk assessment. Stoch Environ Res Risk Assess 26:451. https://doi.org/10.1007/s00477-011-0470-z

Liu T, Jiang BF, Niu WK, Ding SJ, Wang LS, Sun DP, Pei YW, Lin Y, Wang JX, Pang B, Wang XJ (2013) Analysis of clinical features and early warning indicators of death from hand, foot and mouth disease in Shandong province. Chin J Prev Med 47(4):333–336

Murakami Y, Hashimoto S, Taniguchi K, Osaka K, Fuchigami H, Nagai M (2004) Evaluation of a method for issuing warnings pre-epidemics and epidemics in Japan by infectious diseases surveillance. J Epidemiol 14(14):33–40

Pallansch M, Roos R (2007) Enteroviruses: polioviruses, coxsackieviruses, echoviruses, and newer enteroviruses. In: Knipe DM, Howley PM, Griffin DE, Martin MA, Lamb RA et al (eds) Fields virology. Lippincott-William & Wilkins, Philadelphia, pp 840–893

Pearl J (1988) Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann, Burlington

Puenpa J, Theamboonlers A, Korkong S, Linsuwanon P, Thongmee C, Chatproedprai S, Poovorawan Y (2011) Molecular characterization and complete genome analysis of human enterovirus 71 and coxsackievirus A16 from children with hand, foot and mouth disease in Thailand during 2008–2011. Arch Virol 156:2007–2013

Ryu WS, Kang B, Hong J, Hwang S, Kim J, Cheon DS (2010) Clinical and etiological characteristics of enterovirus 71-related diseases during a recent 2-year period in Korea. J Clin Microbiol 48:2490–2494

Tsou YL, Lin YW, Shao HY, Yu SL, Wu SR, Lin HY, Liu CC, Huang C, Chong P, Chow YH (2015) Recombinant adenovaccine expressing enterovirus 71-like particles against hand, foot, and mouth disease. PLoS Negl Trop Dis 9(4):e0003692. https://doi.org/10.1371/journal.pntd.0003692

Tu PV, Thao NT, Perera D, Huu TK, Tien NT, Thuong TC, HowOM Cardosa MJ, McMinn PC (2007) Epidemiologic and virologic investigation of hand, foot, and mouth disease, southern Vietnam, 2005. Emerg Infect Dis 13:1733–1741

Xing W, Liao Q, Viboud C et al (2014a) Hand, foot, and mouth disease in China, 2008–2012: an epidemiological study. Lancet Infect Dis 14(4):308–318

Xing WJ, Sun JL, Liu FF, Yu HJ (2014b) National report of analysis of hand foot and mouth disease monitoring data in China, 2008–2014. Infect Dis Rep 2(5):1–10

Xiong TT, Zhou XT, Zhu Y, Li Y, Wu TS, Ma ZC (2013) Value of precaution model of influenza based on moving percentile method. China Trop Med 13(7):822–824

Yang W, Li ZJ, Lan YJ, Wang JF, Ma JQ, Jin LM, Sun Q, Wei Lv, Lai SJ, Liao YL, Hu WB (2011) A nationwide web-based automated system for early detection and rapid response in China. West Pac Surveill Response J 2(1):10

Yang Z, Ye ZH, You AG, Li X, Jia ZY, Yan L, Wang CJ (2014) Exploration of a moving percentile-based disease-warning model for influenza. Mod Prev Med 41(8):1345–1353

Yang H, Hu SL, Deng ZH, Zhang SY, Luo KW (2015) Analysis of epidemic situation of hand-foot-mouth disease in Hunan during 2009–2013. China Trop Med 15(3):301–303

Yu L, Zhou L, Tan L, Jiang H, Wang Y, Wei S, Nie SF (2014) Application of a new hybrid model with seasonal auto-regressive integrated moving average (ARIMA) and nonlinear auto-regressive neural network (NARNN) in forecasting incidence cases of HFMD in Shenzhen, China. PLoS ONE 9(6):e98241