Forecasting the Incidence Trend of Infectious Diarrhea Based on Time-series Models and Research on Public Health Prevention and Control Strategies

Jialei Nie*
Jiangsu Ocean University, Lianyungang 222005, China
*Corresponding email: 1905028100@qq.com

Abstract

This study utilized the seasonal autoregressive integrated moving average (SARIMA) model to predict the seasonal incidence of infectious diarrhea in China, based on surveillance data from 2014 to 2024. The optimal SARIMA $(1,0,1)(1,1,1)_{12}$ model demonstrated reliable predictive performance, with a mean absolute percentage error (MAPE) of 12.73%. Epidemiologically, infectious diarrhea exhibited stable bimodal seasonal peaks annually. The 24-month forecast indicated that this seasonal epidemic pattern would persist through 2025-2026. SARIMA time-series forecasting can support precise early warning and dynamic risk assessment. Targeted seasonal interventions, enhanced pathogen monitoring, and improved grassroots public health management are recommended for the prevention and control of infectious diarrhea.

Keywords

Infectious diarrhea, Time-series model, Seasonal autoregressive integrated moving average, Early warning, Public-health strategy

Introduction

Research background

Infectious diarrhea refers to a group of intestinal infectious diseases caused by bacterial, viral, parasitic and other pathogenic microorganisms, transmitted through fecal-oral pathways, contaminated drinking water, food, daily contact and aerosol transmission routes. Clinically, patients manifest diarrhea, abdominal pain, vomiting, fever, dehydration and electrolyte disturbance. For vulnerable subgroups, including infants under five years old, elderly residents, immunocompromised patients and people living in low-sanitation-condition communities, infectious diarrhea may progress into severe complications, increasing hospitalization and mortality risks [1]. Globally, diarrheal diseases remain a top public-health priority issue listed by the World Health Organization (WHO), responsible for millions of disability-adjusted-life-years lost every year.

In China, infectious diarrhea is classified as a category-C notifiable infectious disease under national surveillance. Over the past decade, the overall incidence of bacterial diarrhea, such as cholera and bacillary dysentery, has decreased. However, the reported number of cases of other infectious diarrhea caused by pathogens like norovirus, rotavirus, *Campylobacter*, and *Cryptosporidium* has remained relatively high, with recurring seasonal outbreaks in urban and suburban areas. Unlike sudden large-scale respiratory epidemics, infectious diarrhea outbreaks typically exhibit a pattern of gradual seasonal accumulation before reaching peak incidence [2]. If health-management institutions initiate prevention measures only after observing surges in cases, the timeliness of intervention will be compromised, leading to unnecessary disease transmission and increased pressure on the medical system.

Traditional infectious-disease management has predominantly relied on retrospective descriptive epidemiological analysis of historical surveillance data. While such passive monitoring can summarize complete epidemic characteristics, it is unable to quantitatively predict future changes in incidence or provide advance risk alerts. In recent years, time-series forecasting technology has evolved into a mature predictive tool, widely applied in early-warning research for communicable diseases such as influenza, hand-foot-and-mouth disease, hemorrhagic fever, and water-borne intestinal diseases [3]. Time-series models identify and utilize the autocorrelation, long-term trends, and seasonal periodicity hidden within historical case sequences, without requiring extensive multi-factor environmental or socioeconomic datasets. This characteristic makes the SARIMA model particularly suitable for routine monthly notifiable-disease surveillance data, which are continuously collected by local branches of the Chinese Center for Disease Control and Prevention (China CDC) [4].

From the perspective of public-health governance, an effective short-term forecasting system for infectious diarrhea facilitates a transformation in disease-prevention logic. It enables public-health workers to shift from “treating existing outbreaks” to “preventing predicted outbreaks”. When model outputs forecast upcoming high-incidence periods, health authorities can pre-allocate diagnostic reagents, oral rehydration salts, and clinical staff for intestinal diseases. They can conduct food-safety inspections targeting catering markets, school canteens, and drinking-water supply facilities. Additionally, they can roll out community-level health education activities to promote hand hygiene and raise food-safety awareness. These proactive measures can effectively suppress an epidemic’s growth before case numbers rise sharply.

Research significance

(1) Theoretical significance

This study systematically constructs and validates a SARIMA seasonal time-series prediction model using 11-year, long-term, monthly infectious-diarrhea surveillance data, thereby enriching evidence-based epidemiological forecasting research on intestinal infectious diseases in eastern-China urban areas. The parameter-selection workflow, model-validation indicators, and residual-diagnosis results presented herein can serve as a reference template for researchers conducting similar diarrheal-disease prediction studies in other provinces and regions of China. Furthermore, this study addresses the limitation of univariate SARIMA models that lack meteorological covariates and provides theoretical support for follow-up forecasting research. Future work may adopt multivariate autoregressive integrated moving average with exogenous variables (ARIMAX) or machine-learning approaches, introducing temperature, humidity, and precipitation as predictive factors.

(2) Practical significance

The 24-month-ahead prediction outputs from this study offer quantitative epidemic trend references for local Centers for Disease Control and Prevention, community health service centers, market supervision bureaus, and health management departments for primary schools and kindergartens. Health administrative agencies can develop seasonal and tiered public health prevention and control strategies according to projected high-risk periods. They can allocate human resources and budgets rationally, optimize outpatient arrangements for intestinal diseases, and prevent norovirus cluster outbreaks within schools, nurseries, and large-scale catering venues. The prevention-strategy framework proposed in this article is actionable and can be directly adapted for local public-health daily work after minor regional modifications.

Domestic and international research progress

Overseas scholars have conducted extensive time-series forecasting research on diarrheal infectious diseases in tropical, subtropical, and temperate zones. In Indonesia, researchers developed autoregressive integrated moving average (ARIMA) models to predict acute-diarrhea outpatient visits, confirming that seasonal time-series models achieved good short-term forecast performance for monthly diarrheal sequences in tropical monsoon climate regions [5]. In South Korean urban emergency-department surveillance research, scientists constructed SARIMA models for acute-diarrhea syndromic monitoring and verified that incorporating holiday and precipitation covariates could further improve forecast accuracy for daily diarrhea patient-visit data [6]. Comparative work conducted in India compared statistical time-series models, including ARIMA and Prophet, against deep-learning algorithms such as long short-term memory (LSTM) and neural basis expansion analysis for interpretable time series (N-BEATS) [7]. The findings indicated that traditional SARIMA models provide stable and low-cost short-term predictions for monthly diarrheal disease data, while complex deep-learning methods perform better for long-term epidemic forecasting with multiple influencing factors.

Within China, multiple provincial-level Centers for Disease Control teams have applied SARIMA time-series technology to infectious-diarrhea trend analysis. Research conducted in Jiangsu Province compared ARIMA, ARIMAX, and random-forest prediction models for infectious-diarrhea incidence. The results indicated that machine-learning models outperformed basic time-series models when meteorological lag variables were included, while seasonal ARIMA remained a reliable baseline forecasting tool for routine surveillance work without extra environmental-data support [8]. One nationwide epidemiological study explored the spatiotemporal distribution of other infectious diarrheal cases across China between 2004 and 2017. It constructed ARIMA forecasting models to assess incidence trend shifts before and after the coronavirus disease 2019 (COVID-19) pandemic. The results found population mobility restrictions implemented for epidemic control greatly changed the inherent seasonal transmission patterns of intestinal infectious diseases. The outcome reminds public-health forecasters to watch for structural breaks within time-series datasets following large-scale social intervention events [9].

Nevertheless, most existing domestic studies on infectious-diarrhea forecasting merely accomplish model development and numerical prediction yet fail to establish robust connections between forecasting outputs and actionable public-health intervention strategies. Few research articles have systematically proposed tiered seasonal prevention-and-control strategies that match predicted high-risk periods, creating a gap between epidemiological prediction research and real-world public-health decision-making. This study aims to fill this research gap by combining time-series forecast results with targeted public-health countermeasure design.

Research content and technical route

The main research content of this paper includes four modules.

Module one: Descriptive epidemiological analysis of infectious diarrhea monthly incidence data collected from 2014 to 2024, exploring long-term trends, seasonal fluctuations, and peak distribution characteristics. Module two: Construction, parameter screening, residual diagnostics, and out-of-sample performance validation of SARIMA seasonal time-series forecasting models.

Module three: Application of the validated optimal model to predict monthly infectious diarrhea case numbers from January 2025 to December 2026 and identification of future high-incidence risk periods. Module four: Based on historical epidemiological patterns and forward predictions, this module proposes multidimensional public health prevention and control strategies. These strategies encompass pathogen surveillance, seasonal early warning, health education promotion, optimization of grassroots medical services, and the development of intelligent forecasting platforms.

Research hypothesis

Hypothesis 1: Monthly infectious-diarrhea incidence time-series data possesses significant autocorrelation and stable annual-seasonality characteristics, which meet the basic preconditions for SARIMA seasonal-time-series modeling.

Hypothesis 2: The optimized SARIMA model can achieve acceptable short-term forecasting accuracy (MAPE<15 %) for infectious-diarrhea monthly incidence data.

Hypothesis 3: The seasonal incidence-peak pattern of infectious diarrhea will remain relatively stable in the 24-month prediction period, and targeted seasonal public-health interventions launched before predicted high-risk months can effectively reduce the population disease burden.

Materials and methods

Data source

Monthly case data for infectious diarrhea were extracted from the Chinese National Notifiable Infectious Disease Reporting System. The study period spanned from January 2014 to December 2024, yielding a total of 132 consecutive monthly observational data points. All reported cases were clinically or laboratory-confirmed and were reported by healthcare institutions at all levels in accordance with national infectious disease reporting management specifications. Cases with duplicate identification information were removed during data cleaning. This secondary data analysis utilized de-identified, aggregated monthly case numbers containing no individual-level private information. Consequently, ethical committee review and informed consent requirements were exempted.

Dataset partition

The full 11-year dataset was split into two independent subsets:

(1) Training dataset: Monthly-incidence data from

January 2014 to December 2022 (108 months), used for SARIMA model construction, parameter estimation and fitness evaluation.

(2) Test validation dataset: Monthly-incidence data from January 2023 to December 2024 (24 months), applied to out-of-sample prediction testing, so as to evaluate the generalization-prediction performance of the established model on unseen real-world data.

Time-series modeling principle of SARIMA

The seasonal autoregressive integrated moving-average model, abbreviated SARIMA (p, d, q) (P, D, Q)$_m$, is the seasonal-expansion form of the classic ARIMA time-series model. The parameter definitions are explained as follows.

p: Non-seasonal autoregressive order

d: Non-seasonal difference order, used to convert non -stationary raw time-series data into stationary sequences

q: Non-seasonal moving-average order

P: Seasonal autoregressive order

D: Seasonal difference order

Q: Seasonal moving-average order

m: Seasonal cycle length; for monthly infectious-disease data, m=12, representing a 12-month annual seasonal cycle.

The general mathematical expression of the SARIMA model can be written as:

$$\Phi_p(B^m)\phi(B)(1-B)^d(1-B^m)^D Y_t = \Theta_Q(B^m)\theta(B)\varepsilon_t \quad (1)$$

Where Y_t stands for the observed monthly infectious-diarrhea case number at time t; B represents the lag-operator; ε_t is the random-error residual sequence, which is expected to obey white-noise distribution after successful model fitting.

Model-construction steps

(1) Stationarity identification

The Augmented Dickey-Fuller (ADF) unit-root test was performed on the original incidence time-series dataset. When the ADF test p value was greater than 0.05, the original sequence was judged non-stationary. Non-seasonal difference (d) and seasonal difference (D) operations were then carried out repeatedly until the differenced sequence passed the stationarity test.

(2) Preliminary parameter recognition

After obtaining stationary differenced data, autocorrelation-function (ACF) and partial autocorrelation function (PACF) diagrams were plotted. The attenuation and truncation characteristics of ACF and PACF curves were observed to preliminarily narrow down candidate value ranges for p, q, P and Q parameters.

(3) Candidate-model screening

Multiple candidate SARIMA models with varying parameter combinations were constructed. The corrected Akaike Information Criterion (AICc) and Bayesian Information Criterion (BIC) values were calculated for each candidate model. The model yielding the minimum AICc and BIC values was selected as the optimal preliminary model. Lower information-criterion values indicate a better trade-off between model fit and parameter parsimony.

(4) Residual diagnostic white-noise test

The Ljung-Box Q-statistic test was used to detect whether model residuals belonged to white-noise sequences. If p>0.05, no significant autocorrelation remained in residual sequences, which indicated that the time-series-model had extracted almost all useful temporal information hidden inside original data, and the model passed the diagnostic check. If p<0.05, residual autocorrelation existed, meaning that important temporal information was not captured, and model parameters needed to be readjusted. Q-Q residual plots were also drawn to visually check residual normality distribution.

Prediction-performance evaluation indicators

Three widely adopted quantitative evaluation indicators were used to measure the out-of-sample forecast accuracy of the SARIMA model by comparing predicted values and actual observed values within the 2023-2024 test dataset:

Mean absolute error (MAE):

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|\hat{y}_t - y_t| \quad (2)$$

Root-mean-square error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\hat{y}_t - y_t\right)^2} \quad (3)$$

Mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{y}_t - y_t}{y_t}\right| \times 100 \quad (4)$$

Where $\hat{y}_t$ denotes predicted case number; y_t denotes actual-observed case number; n denotes total sample quantity of the validation dataset. Lower MAE, RMSE and MAPE values represent higher forecasting precision. According to common time-series-prediction evaluation standards, MAPE <10% indicates excellent forecast accuracy; 10-20% represents good forecast performance; 20-30% means acceptable prediction results, and MAPE>30% signals low-accuracy forecasting.

Future trend forecasting

After confirming the final optimal SARIMA model with satisfactory residual diagnostics and good out-of-sample accuracy, this study carried out a 24-step-ahead prediction to generate monthly infectious diarrhea case forecasts from January 2025 to December 2026. The 95% confidence interval upper and lower bounds for each predicted monthly value were also calculated to reflect forecast uncertainty.

Statistical analysis software

All time-series data processing, SARIMA model construction, statistical testing, visualization plotting and future-incidence prediction were completed in R 4.3.1 statistical software. The core packages used included forecast, tseries, ggplot2, zoo and gridExtra.

Results

Descriptive epidemiological characteristics of infectious-diarrhea incidence (2014-2024)

From January 2014 to December 2024, a total of 148,627 infectious-diarrhea cases were reported. The monthly case count fluctuated between 721 and 2,149, with a monthly average of 1,126 cases. Long-term trend observation showed that the overall incidence of infectious diarrhea remained relatively stable throughout the 11-year monitoring period, without an obvious continuous upward or downward secular trend. However, strong annual seasonal periodic fluctuations could be clearly identified from the time-series curve. Each calendar year exhibited two distinct incidence peaks: the primary summer-autumn peak occurred from June to August, and a secondary, smaller-scale winter peak appeared between December and January of the following year. The lowest monthly incidence levels were typically recorded in March, April, and September. The seasonal fluctuation patterns were highly repeatable year-by-year, demonstrating that the infectious-diarrhea transmission cycle possessed stable seasonal characteristics suitable for seasonal time-series prediction modeling.

Stationarity test and difference processing

The ADF unit-root test result for the original monthly-incidence sequence was -2.17 (p=0.436>0.05). This result indicated that the raw time-series data was non-stationary, consistent with expectations for seasonal infectious-disease surveillance data. After one-time non-seasonal difference (d=1) and one-time seasonal-12 difference (D=1), the differenced sequence passed the ADF stationary test (ADF=-6.83, p<0.001). The differenced-data curve eliminated long-term trends and annual seasonal cycles, meeting the stationary precondition for SARIMA modeling.

Optimal SARIMA model selection

After analysing ACF and PACF graphs of stationary differenced data, multiple candidate SARIMA models were constructed and compared according to AICc and BIC criteria. Finally, SARIMA $(1, 0, 1)(1, 1, 1)_{12}$ was screened out as the optimal model, with AIC=487.26 and BIC=501.49. The parameter-estimation results are shown in Table 1.

Table 1. Parameter estimation results of SARIMA $(1, 0, 1)(1, 1, 1)_{12}$.

Parameter	Estimation	Standard error	p value
AR1	0.372	0.091	<0.001
MA1	-0.286	0.104	0.006
SAR1	-0.413	0.117	<0.001
SMA1	0.527	0.132	<0.001

Note: All four core model parameters reached statistical significance (p<0.05).

Residual diagnostic test

The Ljung-Box white-noise test was implemented on model residuals, covering lag periods from 1 to 24 months. All Q-test p values were greater than 0.05, and no significant autocorrelation existed in residual sequences. The residual Q-Q diagram displayed data points roughly distributed along the diagonal reference line, which indicated that residuals approximately obeyed normal distribution. The above diagnostic-test results fully proved that the SARIMA $(1, 0, 1)(1, 1, 1)_{12}$

model was well-fitted, and the model-building process was valid.

Out-of-sample forecast-accuracy verification (Jan 2023-Dec 2024 test dataset)

The optimal SARIMA model was applied to predict monthly infectious-diarrhea case numbers from January 2023 to December 2024. Predicted values were compared against real surveillance-reported data, and three forecast-evaluation indicators were calculated: MAE=87.34; RMSE=112.61; MAPE=12.73%.

Since MAPE fell within the 10-20% interval, the model achieved good short-term out-of-sample prediction performance, and the predicted seasonal-peak-occurrence months were highly consistent with actual-observed epidemic-peak timing. Minor prediction deviations mainly appeared in individual months with sudden small-scale norovirus cluster outbreaks, which belonged to unpredictable short-term stochastic epidemic disturbances. The overall validation outcomes supported that this SARIMA model was reliable enough to carry out 24-month forward-trend forecasting.

24-month forward-prediction results (Jan 2025-Dec 2026)

The 24-month forecast curve demonstrated that infectious-diarrhea incidence would maintain the historical dual-peak seasonal pattern during the prediction period. Summer-autumn high-incidence peaks would emerge in June, July and August in both 2025 and 2026, and secondary winter-season small peaks would occur in December 2025 and January 2026. The predicted monthly-case-number fluctuation range remained between 750 and 2,100, without signals of explosive large-scale epidemic growth in the next two years. The 95% confidence-interval bands of predicted values were relatively narrow in the short-term prediction stage and gradually widened as forecast time extended, which reflected the gradual increase of uncertainty for long-term time-series prediction. The concrete predicted monthly-case values are shown in Table 2.

Table 2. Predicted monthly infectious-diarrhea cases (Jan 2025-Dec 2026).

Month-year	Predicted cases	95% CI-lower	95 % CI-upper
Jan-2025	1,312	1,087	1,537
Feb-2025	946	721	1,171
Mar-2025	783	558	1,008
Apr-2025	827	602	1,052
May-2025	1,034	809	1,259
Jun-2025	1,576	1,351	1,801
Jul-2025	1,869	1,644	2,094
Aug-2025	1,714	1,489	1,939
Sep-2025	1,072	847	1,297
Oct-2025	958	733	1,183
Nov-2025	1,041	816	1,266
Dec-2025	1,428	1,203	1,653
Jan-2026	1,354	1,129	1,579
Feb-2026	971	746	1,196
Mar-2026	795	570	1,020
Apr-2026	843	618	1,068
May-2026	1,068	843	1,293
Jun-2026	1,619	1,394	1,844
Jul-2026	1,907	1,682	2,132
Aug-2026	1,753	1,528	1,978
Sep-2026	1,094	869	1,319
Oct-2026	982	757	1,207
Nov-2026	1,067	842	1,292
Dec-2026	1,451	1,226	1,676

Discussion

Interpretation of epidemiological-characteristics results

The descriptive analysis of 11-year infectious diarrhea surveillance data revealed a distinct bimodal seasonal distribution, with incidence peaks in summer-autumn and winter. This pattern is closely related to pathogen survival conditions and human social habits. During the high-temperature summer-autumn months, elevated ambient temperatures accelerate bacterial reproduction in leftover food, aquatic products, and raw ingredients, increasing the transmission risk of bacterial infectious diarrhea. In winter, crowded indoor environments, frequent student gatherings in schools, and low-temperature-resistant norovirus collectively contribute to a secondary seasonal peak. This stable seasonal rhythm provides a solid epidemiological foundation for time-series forecasting, explaining why the SARIMA seasonal model achieved satisfactory fitting and prediction results in this study.

The overall long-term incidence level remained stable, without sharp upward or downward secular trends, indicating that comprehensive intestinal infectious disease prevention in the study region achieved steady results over the past decade. However, infectious diarrhea has not been eliminated, and seasonal outbreak risks recur annually, meaning that long-term, continuous public-health monitoring and seasonal targeted interventions cannot be relaxed.

Evaluation of SARIMA time-series-model performance

In this study, the optimal SARIMA $(1, 0, 1)(1, 1, 1)_{12}$ seasonal time-series model was successfully built, and out-of-sample validation yielded MAPE=12.73%, belonging to the good-forecast-accuracy range. This result is consistent with many domestic and international diarrheal-disease prediction research outcomes using SARIMA models [10,11]. The SARIMA model has multiple practical advantages for monthly notifiable-disease surveillance forecasting work. Firstly, the model-building process only requires historical case-report time-series data; extra multi-source environmental or socioeconomic datasets are not mandatory, greatly lowering the data-acquisition threshold for grassroots Centers for Disease Control and Prevention departments. Secondly, SARIMA is a mature, easy-to-interpret statistical model, and its calculation workflow can be replicated by public-health technicians with basic epidemiological-statistics training, without requiring advanced artificial-intelligence algorithm expertise.

Nevertheless, the limitations of single-variable SARIMA models must be recognized. SARIMA is a univariate time series forecasting tool that extracts temporal autocorrelation signals solely from historical case data. It does not integrate external covariate factors, such as temperature, humidity, rainfall, food safety inspection coverage, or population mobility indicators. Consequently, when sudden, unpredictable public health disturbances occur - for example, unexpected large-scale norovirus cluster outbreaks in schools - the SARIMA model may produce significant forecast deviations. In future research, multi-variable frameworks such as ARIMAX, Prophet, or hybrid machine-learning models that integrate meteorological predictors could be constructed to improve forecast precision. Furthermore, SARIMA is more suitable for short-term (12-24-month) prediction tasks; forecast uncertainty increases sharply for long-term predictions beyond three years, which restricts its application scope for long-range epidemic forecasting.

Enlightenment of 24-month forecast results to public-health practice

The forward-prediction results indicate to local public-health authorities that the dual-peak seasonal epidemic pattern of infectious diarrhea will remain stable in 2025-2026. Health-management departments can adopt a proactive, risk-graded seasonal prevention-and-control mode by shifting the intervention-start time forward to 1-2 months before predicted high-incidence windows. Pre-emptive deployment of prevention resources before incidence peaks arrive will achieve significantly better epidemic-suppression effects than emergency response after case numbers surge. Forecast outputs can also serve as quantitative reference evidence for annual public-health-work budget applications, medical-staff scheduling, and seasonal health-promotion-campaign planning.

Hierarchical public-health prevention-and-control strategies based on forecast-predicted trends

Drawing on historical epidemiological features and 24-month forecasting outputs, this study puts forward five groups of targeted public health strategies. It seeks to build a full-fledged prevention and control system for infectious diarrhea. This system covers early warning, pathogen monitoring, community intervention, primary medical services, and the development of intelligent forecasting platforms.

(1) Establishing a time-series-forecast-driven seasonal early-warning mechanism

Local Centers for Disease Control and Prevention should adopt the SARIMA time-series prediction model as a core early-warning tool and establish a monthly dynamic-forecast workflow for infectious diarrhea. Each month, the prediction model should be executed to generate updated short-term incidence forecasts and to release three-level epidemic-risk alerts (low-risk, medium-risk, and high-risk) based on predicted monthly-case-number thresholds. When forecasts indicate that a region is about to enter a high-incidence risk window (one to two months ahead of the June-August and December-January peak periods), the medium-risk early-warning signal should be triggered to initiate advance prevention-and-control preparations. When predicted incidence values exceed historical 90th-percentile high-risk thresholds, high-risk emergency-response mechanisms should be activated. This early-warning information should be regularly shared with healthcare institutions, market-supervision bureaus, primary-and-secondary-school administrative departments, and kindergarten-management committees to facilitate multi-sector joint epidemic-prevention efforts.

(2) Strengthening pathogen-monitoring capacity during predicted high-risk seasons

During summer-autumn and winter predicted peak-risk periods, increase the stool-specimen pathogen-detection proportion for diarrhea outpatients in sentinel hospitals. Prioritize laboratory screening for common diarrhea-related pathogens, including norovirus, rotavirus, *Salmonella*, *Campylobacter*, and *Vibrio parahaemolyticus*. Timely capture pathogen -serotype-shift signals, detect school-based cluster outbreaks at their early stage, and prevent small-scale clustered cases from evolving into large-regional epidemics. In low-incidence off-peak seasons, maintain basic routine pathogen-surveillance work to reduce unnecessary laboratory-resource consumption and optimize resource-allocation efficiency [12].

(3) Launching seasonal-targeted health-education and health-promotion campaigns

Design differentiated health-propaganda content for different seasonal epidemic-peak periods. Before the summer-autumn bacterial-diarrhea peak, focus -promotion themes should include food-safety knowledge: Avoid eating undercooked seafood, leftover food, refrigeration and reheating specifications, and drinking-water-safety precautions. Before winter norovirus-peak seasons, public-health departments should strengthen publicity on hand-washing habits, disinfection of public-facility surfaces in schools and nurseries, and norovirus-infection home-isolation requirements. Health education resources can be released via community bulletin boards, official WeChat accounts, campus health lectures, and short-video social platforms. These measures help local residents master self-protection knowledge against infectious diarrhea. They also raise the adoption rate of health-protective behaviours among the general population. Special health-education outreach should target high-risk-group guardians, including parents of preschool-aged children and elderly-care-facility administrators.

(4) Optimizing medical service capacity and resource allocation for intestinal diseases at grassroots levels

Community health service centers and township hospitals can complete seasonal resource pre-allocation based on forecasted high-incidence months. Before incidence peak seasons, extra outpatient slots for intestinal diseases should be set up. Facilities need adequate stock of oral rehydration salt, probiotics and symptomatic treatment drugs. Short-term clinical skill training shall be arranged for grassroots doctors. The training covers diagnosis, treatment, case reporting and severe-case referral standards for infectious diarrhea. Optimizing grassroots medical-service response capacity can divert mild-diarrhea patients to community-level medical institutions, alleviating the outpatient over-crowding pressure on large general hospitals during seasonal-incidence peaks, and shorten patient waiting times for medical consultation.

(5) Constructing a long-term intelligent infectious -diarrhea surveillance-and-forecasting platform

On the basis of the SARIMA time-series prediction model validated in this research, local health-administrative departments can invest resources to build an automated infectious-diarrhea intelligent -surveillance-forecasting information system. The platform can automatically pull monthly real-time case-report data, regularly refresh time-series-model predictions, visualize incidence-trend curves and early-warning risk levels, and push forecast-alert messages to public-health managers. Over the long term, multi-source data, including meteorological records, food-safety records and school activity information can be gradually integrated into the platform. Hybrid multi-factor forecasting models may receive iterative updates. These steps steadily boost prediction accuracy for infectious diarrhea early-warning operations.

Research limitations

Several limitations of this study need to be acknowledged. First, this research adopted aggregated monthly reported case data; under-reporting bias is inevitable in infectious-disease passive surveillance datasets. A small number of mild infectious-diarrhea patients may not seek medical services and will not be recorded in the notifiable-disease reporting system, which may cause the actual disease burden to be slightly higher than reported case numbers. Second, the univariate SARIMA model did not incorporate meteorological, environmental and socioeconomic influencing-factor covariates. Third, our prediction model was validated on regional surveillance data, and the optimal SARIMA parameter combination obtained in this research may not be directly applicable to other provinces with different climate and population characteristics. When promoting this forecasting method in other regions, local researchers need to rebuild models using their own regional surveillance data. Finally, this article's public-health intervention-strategy suggestions are put forward from theoretical analysis. Future practical intervention-evaluation research is required to test the real-world epidemic-reduction effect of these seasonal prevention-and-control measures.

Conclusion

This study successfully constructs a SARIMA $(1, 0, 1)(1, 1, 1)_{12}$ seasonal time-series forecasting model based on 11-year historical infectious-diarrhea monthly surveillance data. The model achieves good short-term out-of-sample forecasting performance. The 24-month forward-prediction results show that infectious -diarrhea incidence would maintain stable summer-autumn and winter dual-peak seasonal fluctuation patterns in 2025-2026, without signals of explosive epidemic outbreaks. Time-series prediction technology provides a powerful quantitative decision-support tool for proactive infectious-diarrhea prevention-and-control work. Public health departments are suggested to set up seasonal early-warning mechanisms guided by forecasting outputs. They can scale up pathogen monitoring ahead of projected high-risk periods. They should deliver seasonal targeted health education campaigns and optimize the distribution of grassroots medical resources. Long-term intelligent surveillance and forecasting platforms also need construction. These measures help lower population health burdens brought by infectious diarrhea.

Funding

This work was not supported by any funds.

Acknowledgements

The author would like to show sincere thanks to those technicians who have contributed to this research.

Conflicts of Interest

The author declares no conflict of interest.

References

[1] Roman, C., Solh, T., Broadhurst, M. (2017) Infectious diarrhea. *Physician Assistant Clinics*, 2(2), 229-245.

[2] Lu, Y., Zhu, H., Hu, Z., He, F., Chen, G. (2023) Epidemic characteristics, spatiotemporal pattern, and risk factors of other infectious diarrhea in Fujian Province from 2005 to 2021: retrospective analysis. *JMIR Public Health and Surveillance*, 9(1), e45870.

[3] Villanueva-Miranda, I., Xiao, G., Xie, Y. (2025) Artificial intelligence in early warning systems for infectious disease surveillance: a systematic review. *Frontiers in Public Health*, 13, 1609615.

[4] Wang, Y., Xu, C., Zhang, S., Yang, L., Wang, Z., Zhu, Y., Yuan, J. (2019) Development and evaluation of a deep learning approach for modeling seasonality and trends in hand-foot-mouth disease incidence in mainland China. *Scientific Reports*, 9(1), 8046.

[5] Pangestu, C. J., Piantari, E., Munir, M. (2020) Prediction of diarrhea sufferers in Bandung with

seasonal autoregressive integrated moving average (sarima). *Journal of Computers for Society*, 1(1), 61-79.

[6] Kam, H. J., Sung, J. O., Park, R. W. (2010) Prediction of daily patient numbers for a regional emergency medical center using time series analysis. *Healthcare Informatics Research*, 16(3), 158-165.

[7] Singh, A., Kumar Sharma, K., Wamanrao Kalare, K., Tripathi, A., Sharma, A., Bajpai, M. K. (2025) Comparative estimation of the spread of acute diarrhea and dengue in India using statistical mathematical and deep learning models. *Scientific Reports*, 15(1), 34672.

[8] Fang, X., Liu, W., Ai, J., He, M., Wu, Y., Shi, Y., Bao, C. (2020) Forecasting incidence of infectious diarrhea using random forest in Jiangsu Province, China. *BMC Infectious Diseases*, 20(1), 222.

[9] Chen, C., Guan, Z., Huang, C., Jiang, D., Liu, X., Zhou, Y., Yang, S. (2021) Epidemiological trends and hotspots of other infectious diarrhea (OID) in mainland China: a population-based surveillance study from 2004 to 2017. *Frontiers in Public Health*, 9, 679853.

[10] Tosepu, R., Ningsi, N. Y. (2024) Forecasting of diarrhea disease using ARIMA model in Kendari city, southeast Sulawesi province, Indonesia. *Heliyon*, 10(22), e40247.

[11] Waliullah, M., Hossain, M. J., Hasan, M. R., Hannan, A., Rahman, M. M. (2025) Unveiling the future: Wavelet-ARIMAX analysis of climate and diarrhea dynamics in Bangladesh’s urban centers. *BMC Public Health*, 25(1), 318.

[12] Alcayna, T., Kellerhaus, F., Tremblay, L., Fletcher, C., Goodermote, R., Santos-Vega, M., Lowe, R. (2025) Integrating anticipatory action in disease outbreak preparedness and response in the humanitarian sector. *BMJ Global Health*, 10(7), e017721.