WEP Warwick Evans Publishing

International Journal of Public Health and Medical Research

ISSN: 3006-7707 (Print), ISSN: 3006-3752 (Online) | Volume 3, Number 2, Year 2025
DOI: https://doi.org/10.62051/ijphmr.v3n2.03
Journal homepage: https://wepub.org/index.php/IJPHMR/index

Forecasting Analysis of the Number of AIDS Incidence in China Based on the SARIMA-SVM Combination Model

Mengyao Du [1], Xiao Chang [2]

[1] College of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu 233030, China

[2] Institute of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu 233030, China

ABSTRACT

Based on the monthly incidence data of AIDS, a SARIMA model and a SARIMA-SVM combination model are constructed to study and predict the number of cases, providing a scientific basis for the prevention and control of AIDS. The monthly incidence count data of AIDS from January 2004 to September 2023 are used as the training set, and the monthly incidence count data from October 2023 to September 2024 are used as the test set. Using R software, these two models are applied to predict the monthly incidence count of AIDS in China, and the predictions are compared with the actual values to evaluate the forecasting effectiveness of the single SARIMA model and the SARIMA-SVM hybrid model. The overall trend of AIDS incidence in China from 2004 to 2024 shows an upward trend, with certain seasonal characteristics. The root mean square error (RMSE) of the SARIMA model's forecast results is approximately 578.718, and the mean absolute percentage error (MAPE) is about 11.163%. The root mean square error (RMSE) of the SARIMA-SVM hybrid model's forecast results is approximately 45.643, and the mean absolute percentage error (MAPE) is about 0.838%. The SARIMA-SVM combination model has a good fit and prediction effect on the number of AIDS monthly incidence, and the prediction accuracy is high, which can be used to predict the month-by-month incidence data of AIDS.

KEYWORDS

AIDS incidence number; SARIMA model; SARIMA-SVM combination model; Empirical analysis; Prediction

1. INTRODUCTION

AIDS is not only a global health issue but also a major test for public health systems. The disease is caused by the human immunodeficiency virus (HIV), which can damage the human immune system, making patients vulnerable to various diseases. Although medical research has made progress in the field of AIDS, effective preventive vaccines and targeted cure drugs have not yet been developed. For this reason, AIDS is also known as an “incurable disease”. Once infected with AIDS, it is impossible to return to a healthy state of health, and only a few people in the world have been cured so far.

On July 13, 2023, the Joint United Nations Programme on HIV/AIDS (UNAIDS) released a new report titled "The Road to End AIDS" in Geneva, Switzerland [1]. The report shows that there are currently 39 million people living with HIV worldwide, of which 29.8 million are receiving antiretroviral therapy. In 2022, there are 1.3 million new HIV infections and 630,000 deaths from AIDS-related illnesses.

Content from this work may be used under the terms of CC BY-NC 4.0 licence (https://creativecommons.org/licenses/by-nc/4.0/).
Published by Warwick Evans Publishing.

In the press conference held by the Information Office of the State Council of China on November 1, 2023, Wang Hesheng, the deputy director of the National Health Commission and the head of the National Disease Prevention and Control Administration, underlined that the major infectious diseases currently posing significant threats to the health of the Chinese people mainly encompass AIDS, tuberculosis, hepatitis, and parasitic diseases, etc. [2]. Over the past decade, the mortality rate of AIDS has consistently topped the list among the Class A and B infectious diseases that are legally reported.

AIDS is a complex global problem that has an enormous impact on global health and requires a multifaceted effort to prevent, treat and ultimately control the disease. Only by possessing the ability to scientifically predict the incidence of AIDS can we timely formulate scientific plans and adopt corresponding protective measures and intervention approaches.

Using descriptive epidemiological research methods, Yuxiu Guo conducted an in-depth analysis of the incidence of HIV in China between 2004 and 2016 and explored its epidemiological characteristics [3]. Shuai Wang and Feiran Wei. et al. compare the application of SARIMA model and Holt-Winters model in predicting the number of cases of Sjögren's syndrome [4]; W. W. Wu, Q. Li and D.C. Tian. et al. develop a forecasting model of the incidence of scarlet fever using a seasonal autoregressive integrated moving average (SARIMA) model [5]; Jing Cong, Mengmeng Ren. et al. use SARIMA model to predict the influenza changes [6]; Tohidinik, H.R., Mohebali, M. et al. use the SARIMA model to predict the occurrence of zoonotic cutaneous leishmaniasis (ZCL) and evaluate the effect of climatic variables on disease incidence in the east of Fars province [7]. Yaya Yang, Jie Liang, Shukun Liu used the incidence data of AIDS in Shaanxi Province from 2009 to 2017 to build a GM (1, 1) model to predict the trend of AIDS incidence [8]. Shunyong Li. et al. constructed ARIMA and Prophet models to predict the number of AIDS cases and their trends in China [9]. Chen Kui, Chenxue Dong et al. Predicting the number of AIDS cases in China by building an autoregressive integral sliding average (ARIMA) model [10]. Md Abu Sayeed, Azizur Rahman. et al. employ advanced interpretable ML techniques to anticipate and understand the factors affecting infant mortality in Bangladesh, overcoming the shortcomings of the conventional logistic regression (LR) model [11]. Zeming Li and Yanning Li explore models suitable for the incidence of AIDS in China, aiming to establish a professional and feasible disease prediction model for the prevention and control of AIDS [12].

Zaijin Guo. et al. predicted the number of reported AIDS cases in Jiangsu Province in the first half of 2022 by constructing a combined ARIMA-SVM model [13]; Malan et al. explored and compared the application of the SARIMA model and the Holt-Winters model in the prediction of AIDS cases in the Xinjiang Production and Construction Corps, which provided a scientific basis for the prevention and control of AIDS [14]. The overall trend of HIV/AIDS epidemic in China is at a low prevalence level, but the infection rate among specific groups such as the elderly and young students has increased [15].

Despite the many challenges, the path to ending AIDS is clear. Therefore, it is important to explore and study the prediction model of the number of AIDS incidence, which can help to better carry out the prevention of AIDS (AIDS) and effectively improve the level of AIDS protection and treatment. In this study, the SARIMA single model and the SARIMA-SVM combined model were used to comprehensively study and predict the monthly incidence data of AIDS in China over the past two decades, with a view to providing a scientific and effective method for predicting the incidence of AIDS in the future.

2. CONSTRUCTION AND APPLICATION OF MODEL

2.1. Introduction to Modeling Theory

2.1.1. SARIMA model

The SARIMA model, Seasonal Autoregressive Integrated Moving Average Model, is a widely used forecasting model in time series analysis. It combines the non-seasonal and seasonal components of the ARIMA model for time series data with seasonal patterns.The SARIMA model can be expressed as:

$$\Phi_P(B^m)\phi_p(B)\nabla^d\nabla_m^D y_t = \mu + \Theta_Q(B^m)\theta_q(B)\epsilon_t.$$

Its general form is SARIMA (p, d, q) (P, D, Q, s), where p, d, q are the orders of the non-seasonal component, corresponding to autoregressive, difference and moving average, respectively; P, D, Q are the orders of the seasonal component, again corresponding to autoregressive, seasonal difference and seasonal moving average, respectively. s is the length of the seasonal cycle.

2.1.2. SVM model

Support Vector Machine (SVM) is a widely used supervised learning technique originally designed for data classification and regression prediction, and some scholars have extended it to time series analysis. It has good generalization ability and performs well in dealing with nonlinear problems, small sample data, and high-dimensional data recognition [16]. The core principle is to transform the data to a high-dimensional feature space using kernel tricks and subsequently perform a linear partition within this space with the introduction of penalty terms to transform the regression problem into an optimization problem [16].The SVM model determines the parameter penalty factors C and γ according to the principle of minimizing the error.

Suppose there are k training sets $\{(x_i, y_i), i = 1,2,\cdots,k\}$,where $x_i \in R^D$ ($x_i$contains D features), $i = 1,2,\cdots n$, and n is the n D-dimensional vectors $y_i \in R$, $F = \{f|R^D \to R\}$. For the nonlinear regression task, linear regression can be performed by applying a nonlinear function to transform the original data points into a higher dimensional feature space F[17]. The training sample x_i satisfies the condition:

$$y_i(x_i \cdot \omega + b) - 1 + \xi >> 0, \quad \xi >> 0.$$

ξ_i is called the slack variable and represents a deviation of a given training sample with respect to the optimal classification surface. And $\xi_i \geq 0, i = 1,2,\cdots,n$, ω is an n-dimensional weight, i.e. $\{\omega_1, \omega_2, \cdots, \omega_n\}$. b is a single number called deviation. The regression problem can be equated to solving the following optimization problem:

$$\begin{cases} min\ \frac{1}{2}||\omega||^2 + c\sum_{i=1}^{n}\xi_i, \\ \text{s.t.}\, y_i(x_i \cdot \omega + b) - 1 + \xi \geq 0, \quad i = 1,2,\cdots,n, \end{cases}$$

Where c is the penalty parameter which can be used to balance the size of the slack variables and the classification boundary, $||\omega||$ represents the Euclidean paradigm (distance from the origin to the vector ω), and the solution to the above equation is the optimal final discriminant function $f(x_i) = \text{sign}\left(\sum_{i=1}^{n} a_i y_i K(x_i, x_j) + b\right)$, where $K(x_i, x_j)$ is the kernel function. The choice of kernel function is crucial for constructing support vector machines, and different kernel functions can form their respective SVM models [17]. In this paper, the radial basis (RBF) kernel function is chosen because

the RBF kernel function belongs to the localized kernel function, which has strong nonlinear fitting ability and efficient learning ability [18]. Its expression is: $K(x_m, x_n) = exp\left(-\frac{\|x_m - x_n\|}{2\sigma^2}\right)$, $m = 1,2,\cdots,N$, (σ is the kernel density).

2.1.3. SARIMA-SVM combination model

Guided by the progressive error correction ideology, a SARIMA-SVM combined forecasting model can be constructed. First, a SARIMA model is used to fit and predict the incidence data of AIDS in China, extracting information on linear trends. Then, on the basis of the SARIMA model, an SVM model for the residual sequence is constructed to extract the nonlinear information omitted in the residual sequence of the SARIMA model. By combining the results of the two-step predictions, the final forecast results of the combined model can be obtained.

2.2. Data Pre-processing

The original research data for AIDS required in this article comes from the Chinese Centers for Disease Control and Prevention (CDC) and the National Disease Control and Prevention Administration (NDCPA).

Firstly, the time series plot of the monthly incidence of AIDS from January 2004 to September 2024 is depicted in Fig. 1. Subsequently, the original data series was subjected to trend analysis using the decompose() function in R software, and the results of the decomposition are shown in Fig. 2.

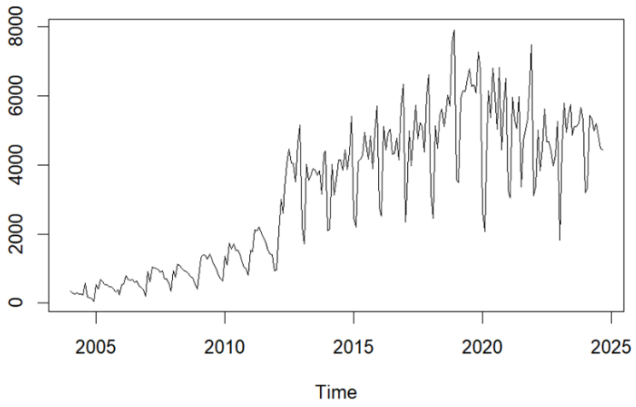

Figure 1. The time series plot of the monthly incidence of AIDS

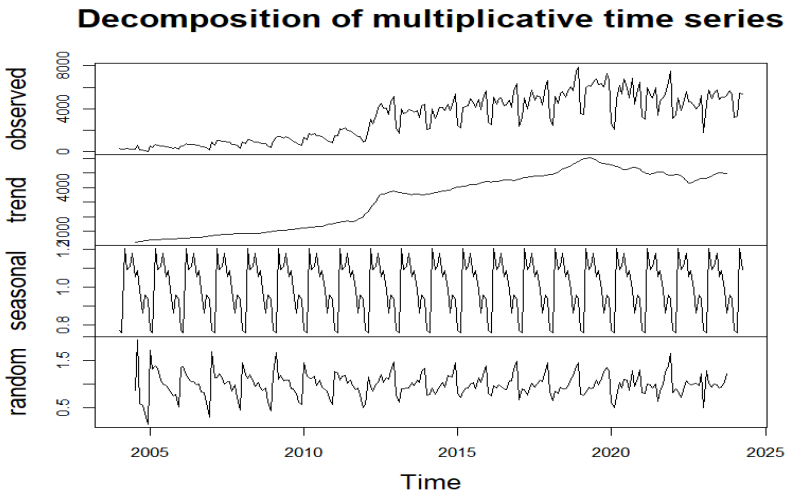

Figure 2. Trend decomposition of the original series

From the above figures, it is evident that the monthly incidence data of AIDS in China exhibit significant seasonal and long-term trends.

An Augmented Dickey-Fuller (ADF) test was conducted on the original data series, and the results are presented in Fig. 3. According to the test results, the p-value is 0.1643, which is higher than the 0.05 confidence level. This means we cannot reject the null hypothesis that the series is a stationary sequence, thus further processing of the sequence is required.

```
        Augmented Dickey-Fuller Test

data:  xts
Dickey-Fuller = -2.9775, Lag order = 6, p-value = 0.1643
alternative hypothesis: stationary
```

Figure 3. ADF test for the original series

First-order differencing was performed on the original data series, and the resulting first-order difference plot is shown in Fig. 4. Additionally, due to the evident seasonality shown in the trend decomposition plot of the original series, we also applied seasonal differencing to the series, and the plot of the series after seasonal differencing is presented in Fig. 5.

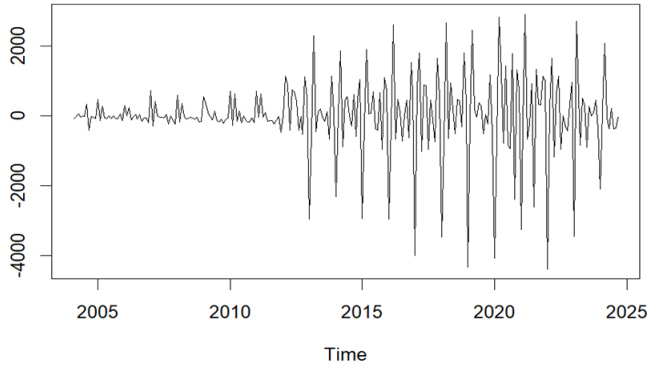

Figure 4. The first-order difference plot

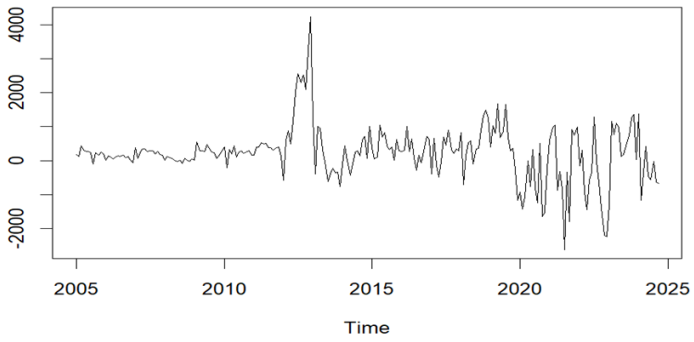

Figure 5. The plot of the series after seasonal differencing

An ADF test was conducted on the series after seasonal differencing, yielding a p-value of 0.01007, which is less than 0.05. At the 0.05 confidence level, we reject the null hypothesis, indicating that the series after seasonal differencing is a stationary sequence. A Ljung-Box test for white noise was also conducted on the seasonally differenced series, resulting in a p-value of approximately 0.000, which is less than 0.05, leading to the rejection of the null hypothesis and indicating that the series is not a white noise sequence.

The original series, after seasonal differencing treatment, has been transformed into a stationary non-white noise sequence, meeting the conditions for further model construction work

2.3. Construction of the SARIMA Model

Based on the original data sequence x_t, we can construct the SARIMA model. The monthly incidence of AIDS in China from January 2004 to September 2024 is divided into a training set and a test set for fitting and forecasting. The forecast results are denoted as $\hat{\ell}_t$.

In Fig. 6 and Fig. 7, we present the autocorrelation (ACF) and partial autocorrelation (PACF) plots of the sequence after seasonal differencing. These plots indicate that the sequence no longer exhibits significant seasonal fluctuations and non-stationarity after differencing. The tailing phenomenon in the ACF and PACF plots suggests that the sequence has been stabilized, providing the conditions necessary for the establishment of the SARIMA model.

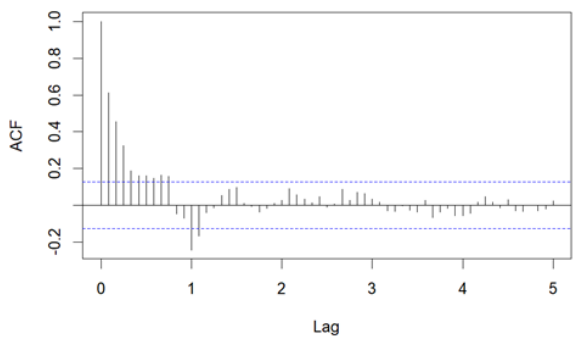

Figure 6. The ACF plot

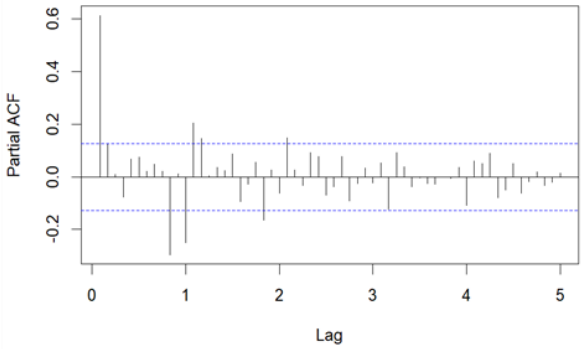

Figure 7. The PACF plot

For the training set data, the optimal SARIMA model is automatically selected based on the principle of the minimum AIC criterion function. The best SARIMA model parameters obtained are (4, 0, 3) (0, 1, 1, 12), with an AIC of 3497.78. Therefore, we construct the SARIMA (4, 0, 3) (0, 1, 1, 12) model.

The residual Q-Q plot shows that the residual sequence generally conforms to the trend of a normal distribution. This indicates that the SARIMA model can effectively fit the data, providing a basis for subsequent forecasting.

The Ljung-Box test for white noise is performed on the residual sequence of the SARIMA (4, 0, 3) (0, 1, 1, 12) model, with a p-value of 0.8897 > 0.05, indicating that there is no significant correlation between the residual sequences. This suggests that the model can be used for fitting and forecasting.

The SARIMA model was applied to fit the AIDS training set data (in person) and the MAE (Mean Absolute Error) of the fitted data was obtained as 332.3126; RMSE (Root Mean Square Error) was 531.2037.

The fitting effect diagram of the SARIMA model is depicted in Fig. 8.

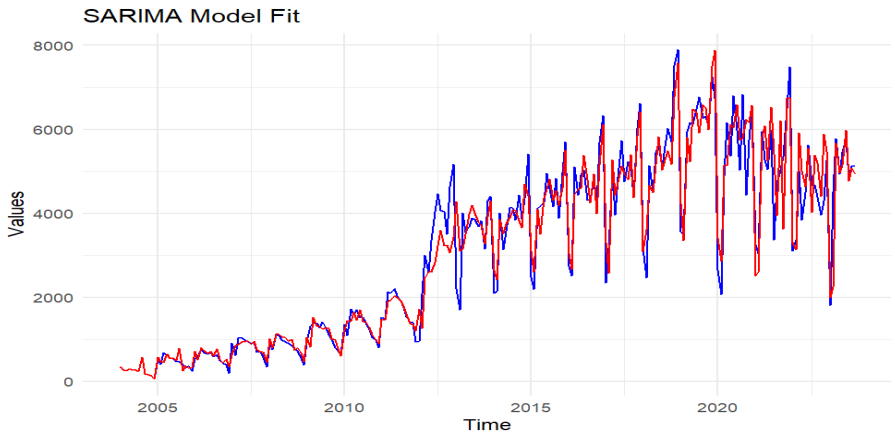

Figure 8. SARIMA model fitting results

2.4. SARIMA Forecast Results

The SARIMA (4, 0, 3) (0, 1, 1, 12) model is used to forecast the monthly incidence number of AIDS from October 2023 to September 2024 in the test set. with the forecast results $\hat{\ell}_t$ presented in Table 1 (retaining three decimal places).

Table 1. The forecast results of SARIMA

date	$\hat{\ell}_t$	date	$\hat{\ell}_t$
2023.10	4929.041	2024.04	5238.775
2023.11	5555.166	2024.05	5382.452
2023.12	6193.985	2024.06	6034.635
2024.01	2822.255	2024.07	5221.764
2024.02	4110.023	2024.08	5183.182
2024.03	5694.327	2024.09	5353.429

Based on the SARIMA model's forecasting results for the monthly incidence of AIDS in China, the average absolute error (MAE) between the actual and forecasted values is calculated to be approximately 503.1301. The maximum and minimum values of the relative error (RE) differ significantly. This indicates that the relative error (RE) between the predicted and actual values is not stable and is somewhat volatile. One reason for the error is that the SARIMA model failed to extract the nonlinear information from the original sequence.

The graph of the SARIMA model's forecasted values (in units of persons) is depicted in Fig. 9 as follows.

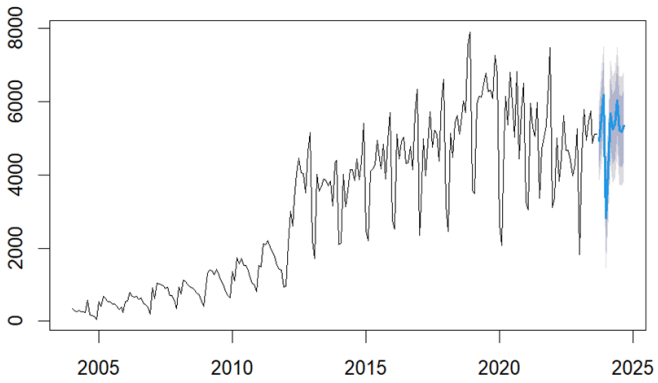

Figure 9. SARIMA model prediction plot

The limitations of the SARIMA model in capturing the nonlinear characteristics of the original data sequence cause fluctuations in its forecasting performance. Therefore, this paper is prepared to further construct an SVM model for the residual sequence of the SARIMA model to extract the omitted nonlinear part, thereby better forecasting the monthly incidence data of AIDS.

2.5. SVM Model of Residual Sequence

On the basis of the data analysis of the original sequence using the SARIMA model, the residual sequence $\{e_t\}$ is constructed as an input vector, which contains the nonlinear relationships not extracted by the SARIMA model from the original sequence [17], and an SVM model is built for this residual sequence. The residual data from January 2004 to September 2024 are fitted and forecasted, with the forecasted results denoted as $\hat{e}_t$. The nonlinear characteristics omitted by the SARIMA model are $(e_t - \hat{e}_t)$.

The residual data is still divided into a training set and a test set, where the training set data are the residual values from the SARIMA model's fitting of the AIDS incidence from January 2004 to September 2023, and the test set data are the residual values from the SARIMA model's forecasting of the AIDS incidence from October 2023 to September 2024.

In the R software, the svm() function is used to automatically determine the parameters of the kernel function, and the Radial Basis Function (RBF) kernel is selected here. The reason is that the RBF kernel is a local kernel function with strong nonlinear fitting capabilities and efficient learning ability.

After multiple fitting tests on the training set residual data, the penalty parameter that minimizes the error is found to be: $C = 1,\ \gamma = 1$. At this point, the minimum error value $\varepsilon = 0.1$. The SVM model for the residual sequence is used to forecast the residual data in the test set. The residual values e_t and their forecasted values $\hat{e}_t$ from October 2023 to September 2024 are presented in Table 2.

Table 2. The residual prediction values of SVM

date	e_t	$\hat{e}_t$	date	e_t	$\hat{e}_t$
2023.10	280.9586	-75.948	2024.04	118.2248	-35.916
2023.11	108.8341	-69.714	2024.05	-395.452	-28.727
2023.12	-898.985	-63.287	2024.06	-833.635	-21.429
2024.01	371.745	-56.680	2024.07	-389.764	-14.036
2024.02	-766.023	-49.906	2024.08	-698.182	-6.563
2024.03	-272.327	-42.980	2024.09	-903.429	0.975

2.6. Construction and Application of the SARIMA-SVM Combination Model

Based on the forecast results $\hat{\ell}_t$ from the SARIMA model, the residual sequence e_t, and the forecasted values $\hat{e}_t$ from the SVM model for the residual sequence, the SARIMA-SVM hybrid model can be constructed. The calculation method for the forecasted values of the SARIMA-SVM combination model is given by $\hat{x}_t = \hat{\ell}_t + (e_t - \hat{e}_t)$.

By integrating the forecast results from the SARIMA model and the SVM model, we obtained the final forecasted values $\hat{x}_t$ for the SARIMA-SVM combination model, which are presented in Table 3.

Table 3. The prediction values of SARIMA-SVM

date	$\hat{x}_t$	date	$\hat{x}_t$
2023.10	5285.947	2024.04	5392.916
2023.11	5733.714	2024.05	5015.727
2023.12	5358.288	2024.06	5222.429
2024.01	3250.68	2024.07	4846.036
2024.02	3393.906	2024.08	4491.563
2024.03	5464.981	2024.09	4449.025

According to the forecasted values given in the above table for the SARIMA-SVM model, the average absolute error (MAE) between the actual and forecasted values is calculated to be approximately 38.84685, which is less than the MAE of the forecasted values from the single SARIMA model mentioned above. This indicates that constructing an SVM model for the residual sequence to extract the nonlinear part not captured by the SARIMA model is of significant importance.

3. EVALUATION OF THE PERFORMANCE OF SARIMA AND SARIMA-SVM

To assess the forecasting performance of the SARIMA model and the SARIMA-SVM combination model, the relative errors (RE) of the forecasted values from both the SARIMA and SARIMA-SVM models are listed separately, retaining four decimal places, as seen in Table 4.

Data from the table below shows that the RE of the SARIMA model fluctuates significantly, with a large difference between the maximum and minimum values, and the mean relative error (MRE) is approximately −0.0764. This indicates that the forecasting effect of the SARIMA model on the monthly incidence of AIDS in China is not stable and has certain limitations. In contrast, the RE of the SARIMA-SVM hybrid model is generally small and does not fluctuate significantly. The calculated MRE for the SARIMA-SVM hybrid model is approximately −0.0083, which is significantly smaller than that of the SARIMA model. Therefore, constructing an SVM model for the residual sequence to extract the nonlinear part not captured by the SARIMA model is of significant importance.

Table 4. Forecast values and RE of SARIMA and SARIMA-SVM

date	Real Value	Forecast Value		RE	
		SARIMA	SARIMA-SVM	SARIMA	SARIMA-SVM
2023.10	5210	4929.041	5285.947	0.0539	-0.0146
2023.11	5664	5555.166	5733.714	0.0192	-0.0123
2023.12	5295	6193.985	5358.288	-0.1698	-0.0120
2024.01	3194	2822.255	3250.68	0.1164	-0.0178
2024.02	3344	4110.023	3393.906	-0.2291	-0.0149
2024.03	5422	5694.327	5464.981	-0.0502	-0.0079
2024.04	5357	5238.775	5392.916	0.0221	-0.0067
2024.05	4987	5382.452	5015.727	-0.0793	-0.0058
2024.06	5201	6034.635	5222.429	-0.1603	-0.0041
2024.07	4832	5221.764	4846.036	-0.0807	-0.0029
2024.08	4485	5183.182	4491.563	-0.1557	-0.0015
2024.09	4450	5353.429	4449.025	-0.2030	0.0002

Additionally, this paper uses the root mean square error (RMSE) and the mean absolute percentage error (MAPE) as indicators to measure the performance of the models [17]. The smaller the values of these two indicators, the better the forecasting effect of the model. The empirical conclusion is that the more the value of MAPE is less than 10%, the higher the model prediction performance is.

The RMSE and MAPE indicator values for both the SARIMA model and the SARIMA-SVM combination model are calculated separately, retaining three decimal places, with the results shown in Table 5.

Table 5. The Forecast accuracy of SARIMA and SARIMA-SVM

Model	RMSE	MAPE [%]
SARIMA	578.718	11.163
SARIMA-SVM	45.643	0.838

The model forecast accuracy table shows that the MAPE value of the SARIMA model is approximately 11.163%, which is greater than 10%; the MAPE value of the SARIMA-SVM model is approximately 0.838%, which is significantly less than 10%. Therefore, it can be determined that the SARIMA-SVM combination model has better predictive performance.

The combined SARIMA-SVM model effectively extracts the linear and nonlinear parts of the month-by-month AIDS incidence data in China, and the accuracy of this model is significantly improved compared with the single model, which can be used as a powerful tool to predict the incidence of AIDS in China.

4. CONCLUSION

This paper selects the data of monthly AIDS incidence from January 2004 to September 2024 in China, divides it into a training set and a test set, and constructs a single SARIMA model and a SARIMA-SVM combination model to fit and predict the monthly AIDS incidence data. It obtains the forecast values and relative errors (RE) of both models and calculates the RMSE and MAPE indicator values. By comparing the predictive accuracy of the two models, it concludes that the SARIMA-SVM hybrid model can achieve better forecast results for AIDS incidence data, and its predictive performance is significantly higher than that of the single model. Therefore, this paper recommends using the SARIMA-SVM model to predict the monthly AIDS incidence data in China.

The SARIMA-SVM combination model combines the time series analysis capabilities of the SARIMA model with the generalization capabilities of the Support Vector Machine (SVM), providing an effective analytical tool for epidemiological research on AIDS and public health decision-making [17]. However, the incidence of AIDS includes various factors, such as medical factors, behavioral factors, socio-economic factors, and cultural factors. These factors contribute to the instability of AIDS incidence data. Therefore, in the process of forecasting and statistical analysis of similar infectious diseases like AIDS, we should keep abreast of their dynamic changes and take effective measures to achieve the desired disease prediction and protection effects.

To enhance the predictive accuracy and stability of the SARIMA-SVM combination model, we must continue to optimize parameters and improve algorithms, and integrate more dimensions of data, such as socio-economic factors and the distribution of medical resources, to enhance the model's explanatory and predictive power. Achieving real-time updates and predictions of the model will allow for rapid response to changes in the epidemic and timely adjustment of control strategies.

Using historical incidence data, methods such as time series analysis and machine learning algorithms are employed to predict the future incidence trends of AIDS. These models play an important role in improving predictive accuracy, assisting public health decision-making, and optimizing resource allocation. Despite challenges in data collection and processing, limitations in model generalization capabilities, and rapidly changing epidemiological environments, ongoing research and technological innovation are promoting the development of this field, providing a scientific basis for the prevention and control strategies of AIDS.

REFERENCES

[1] “The Joint United Nations Programme on HIV/AIDS (UNAIDS) has released a new report: In 2022, there were 1.3 million new HIV infections globally”. National Center for AIDS/STD Control and Prevention, China CDC. 2023. https://ncaids.chinacdc.cn/xxgx/yqbh/

[2] “In the Name of Love, Speaking for 'AIDS' —— A Data-Driven Review of 15 Years of AIDS in China”. The Paper. 2023. https://www.thepaper.cn/newsDetail_forward_25641879

[3] Yuxiu Guo. Analysis of AIDS Epidemic Characteristics and Prediction of Monthly Incidence Rate in China from 2004 to 2016 [D]. Anhui Medical University, 2019.

[4] Wang S, Wei F, Li H, Wang Z, Wei P. Comparison of SARIMA model and Holt-Winters model in predicting the incidence of Sjögren's syndrome. Int J Rheum Dis. 2022 Nov; 25(11):1263-1269. doi: 10.1111/1756-185X.14417. Epub 2022 Aug 12. PMID: 35962522.

[5] Wu WW, Li Q, Tian DC, Zhao H, Xia Y, Xiong Y, Su K, Tang WG, Chen X, Wang J, Qi L. Forecasting the monthly incidence of scarlet fever in Chongqing, China using the SARIMA model. Epidemiol Infect. 2022 Apr 21; 150:e90. doi: 10.1017/S0950268822000693. PMID: 35543101; PMCID: PMC9102071.

[6] Cong J, Ren M, Xie S, Wang P. Predicting Seasonal Influenza Based on SARIMA Model, in Mainland China from 2005 to 2018. International Journal of Environmental Research and Public Health. 2019; 16(23):4760.

[7] Tohidinik, H.R., Mohebali, M., Mansournia, M.A., Niakan Kalhori, S.R., Ali-Akbarpour, M. and Yazdani, K. (2018), Forecasting zoonotic cutaneous leishmaniasis using meteorological factors in eastern Fars province, Iran: a SARIMA analysis. Trop Med Int Health, 23: 860-869.

[8] Yaya Yang, Jie Liang, Shukun Liu. Application of GM (1, 1) Model for Forecasting the Incidence Trend of AIDS in Shaanxi Province [J]. Modern Medicine and Health Research Electronic Journal, 2019, 3(21):141-143.

[9] Shunyong Li, Kexin Li. Application of ARIMA and Prophet Models in Predicting the Number of AIDS Cases [J]. Henan Science, 2020, 38(09):1387-1393.

[10] Kui Chen, Chenxue Dong, Jiayue Lu et al. Forecast of the Number of AIDS Patients in China Based on ARIMA Time Series Model [J]. Primary Health Care, 2022, 36(03):91-93.

[11] Sayeed, M.A., Rahman, A., Rahman, A. et al. On the interpretability of the SVM model for predicting infant mortality in Bangladesh. J Health Popul Nutr 43, 170 (2024).

[12] Li, Z., Li, Y. A comparative study on the prediction of the BP artificial neural network model and the ARIMA model in the incidence of AIDS. BMC Med Inform Decis Mak 20, 143 (2020).

[13] Zaijin Guo, Hao Gong, Yunshui Li et al. Application of ARIMA and ARIMA-SVM hybrid model in the prediction of HIV incidence in Jiangsu [J]. Chinese Preventive Medicine, 2023, 24(08): 857-860. DOI:10.16506/j.1009-6639.2023.08.017.

[14] Lan Ma, Wenwen Yang, Xiaoling Ma et al. Application of SARIMA model and Holt-Winters exponential smoothing model in HIV/AIDS case prediction in Xinjiang Production and Construction Corps [J/OL]. Journal of Preventive Medicine Information:1-7 [2024-06-21]. https://doi.org/10.19971/j.cnki.1006-4028.230532.

[15] Mengmeng Li, Jinghui Zhang. Research Progress on the Status Quo of AIDS Epidemic [J]. Advances in Clinical Medicine, 2023, 13(2): 1601-1606.

[16] Chaoyang Guo, Huiying Liu, Shengchao Liu et al. Application of SARIMA-SVM combination model in the prediction of gonorrhea incidence [J]. Modern Disease Control and Prevention, 2024, 35(02):113-117. DOI:10.13515/j.cnki.hnjpm.1006-8414.2024.02.011.

[17] Yu Zhao, Suisui Chen, Liping Lv. Application of SARIMA-SVM Model in Tuberculosis Incidence [J]. Mathematics in Practice and Theory, 2023, 53(09):133-141.

[18] Fan Zhang, Luning Xun, Jixin Sun et al. Mortality prediction of road traffic injury based on ARIM-SVM combination model [J]. Modern Preventive Medicine, 2021, 48(10):1742-1746.