Infectious Disease Modelling 9 (2024) 224–233

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

Time series models in prediction of severe fever with thrombocytopenia syndrome cases in Shandong province, China

Zixu Wang [a, b, 1], Wenyi Zhang [c, 1], Ting Wu [d, 1], Nianhong Lu [a], Junyu He [e, f], Junhu Wang [a], Jixian Rao [a], Yuan Gu [d], Xianxian Cheng [b], Yuexi Li [a, **], Yong Qi [a, *]

[a] Pest Control Department, Huadong Research Institute for Medicine and Biotechniques, Nanjing, Jiangsu province, 210002, China
[b] Bengbu Medical College, Bengbu, Anhui province, 233030, China
[c] Chinese PLA Center for Disease Control and Prevention, Beijing, 100071, China
[d] Jinling Hospital, Medical School of Nanjing University, Nanjing, Jiangsu province, 210002, China
[e] Ocean College, Zhejiang University, Zhoushan, 316021, China
[f] Ocean Academy, Zhejiang University, Zhoushan, 316021, China

ARTICLE INFO

Article history:
Received 25 October 2023
Received in revised form 19 December 2023
Accepted 11 January 2024
Available online 17 January 2024
Handling editor: Daihai He

Keywords:
Severe fever with thrombocytopenia syndrome
Long short-term memory
Prediction model
Autoregressive integrated moving average
Prophet

ABSTRACT

Severe fever with thrombocytopenia syndrome (SFTS) is an emerging infectious disease caused by the SFTS virus (SFTSV). Predicting the incidence of this disease in advance is crucial for policymakers to develop prevention and control strategies. In this study, we utilized historical incidence data of SFTS (2013–2020) in Shandong Province, China to establish three univariate prediction models based on two time-series forecasting algorithms Autoregressive Integrated Moving Average (ARIMA) and Prophet, as well as a special type of recurrent neural network Long Short-Term Memory (LSTM) algorithm. We then evaluated and compared the performance of these models. All three models demonstrated good predictive capabilities for SFTS cases, with the predicted results closely aligning with the actual cases. Among the models, the LSTM model exhibited the best fitting and prediction performance. It achieved the lowest values for mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE). The number of SFTS cases in the subsequent 5 years in this area were also generated using this model. The LSTM model, being simple and practical, provides valuable information and data for assessing the potential risk of SFTS in advance. This information is crucial for the development of early warning systems and the formulation of effective prevention and control measures for SFTS.

* Corresponding author.
** Corresponding author.
E-mail addresses: liyxi2007@126.com (Y. Li), qslark@126.com (Y. Qi).
Peer review under responsibility of KeAi Communications Co., Ltd.
[1] These authors have contributed equally to this work.

https://doi.org/10.1016/j.idm.2024.01.003

1. Introduction

Severe fever with thrombocytopenia syndrome (SFTS) is an emerging infectious disease, with a mortality of 12%–50% (Li et al., 2018; Yu et al., 2011). It is caused by the SFTS virus (SFTSV), now known as Dabie bandavirus, and classified in the order *Bunyavirales*, family *Phenuiviridae*, and genus *Phlebovirus*. The virus is mainly prevalent in Eastern Asia, including China, Japan, South Korea, etc. (Kato et al., 2016; Shin et al., 2015; Yu et al., 2011). However, a similar virus causing SFTSV-like symptoms has been isolated in the United States, indicating a potentially wide distribution and public health threat of this virus (McMullan et al., 2012). Due to the high threat brought by the virus, the World Health Organization (WHO) listed it as one of the nine most infectious diseases on the priority list in 2017.

Predicting the incidence of this disease can provide important information for prospectively making prevention and control strategies for policymakers (Dharmarajan et al., 2022). For example, using predicted data, local disease prevention and control institutions or hospitals can prepare sufficient personnels and medical supplies, such as detection reagents, therapeutic drugs, disinfection materials, and insecticides. Additionally, during periods of high predicted incidence of SFTS, government departments can issue early warnings to residents and conduct disease prevention education in advance to reduce the actual incidence.

Previous studies have mainly focused on multivariate analysis methods to predict the incidence of SFTS. These studies analyzed various environmental, meteorological, and social factors, among others, to identify incidence-related risk factors and establish prediction models based on these factors (Cho et al., 2021; Deng et al., 2022; Ding et al., 2014; Sun et al., 2018, 2021; Wu et al., 2020). However, few studies have reported models for prediction of SFTS incidence just based on the historical incidence data, which may be more convenient and practical.

As a tick-borne disease, the incidence of SFTS exhibits strong seasonal regularity, with more cases occurring between May and October in China. Based on this seasonal distribution, we established three univariate prediction models using Autoregressive Integrated Moving Average (ARIMA), Prophet, and Long Short-Term Memory (LSTM) algorithms, respectively. These models were developed using historical incidence data from Shandong Province, China, and their prediction performance was further evaluated and compared. The number of SFTS cases in the subsequent 5 years in this area were also predicted using the optimal model to give useful information to the local public health agencies.

2. Materials and methods

2.1. Data collection and preprocessing

We collected daily confirmed SFTS cases in Shandong province from January 2013 to December 2020 from the Public Health Science Data Center (www.phsciencedata.cn/Share/). To ensure optimal training performance for machine learning, we used the Min-Max normalization method to scale and normalize the data with large variations. For the LSTM model, we further transformed the normalized data into LSTM-recognizable format using the create_dataset function.

In general, the data from January 2013 to December 2019 (containing data of 84 months) were used as the training set and those from January to December 2020 (containing data of 12 months) were used as the test set for model construction. For the LSTM model, the data from January 2013 to December 2018 (containing data of 72 months) were used as input, and the data from January to December 2019 (containing data of 12 months) were used as output for training.

2.2. Model construction

2.2.1. ARIMA model

The ARIMA model is a widely used statistical method for time series data prediction and has demonstrated excellent performance in various fields since its introduction. The basic equation for ARIMA is as follows:

$$(1\text{-B})^{\mathrm{d}} Y_t = c + \left(\frac{\theta(B)}{\varphi(B)}\right)\varepsilon_t \tag{1}$$

Where d is the degree of differencing, Y_t is the value of the time series at time t, c is the constant term, ε_t is the white noise sequence, B is the backshift operator, θ(B) is the moving average operator, and φ(B) is the autoregressive operator. However, in this study, due to the obvious seasonality of monthly reported SFTS cases, we adopted the seasonal ARIMA (SARIMA) model. The formula for the SARIMA model is as follows:

$$(1-B)^d\left(1-B^S\right)^D Y_t = c + \left(\frac{\theta(B)\theta_S\left(B^S\right)}{\varphi(B)\varphi_S\left(B^S\right)}\right)\varepsilon_t \tag{2}$$

Where D is the number of seasonal differences, φ(B) and θ(B) are the seasonal autoregressive and moving average operators, respectively.

2.2.2. Prophet model

The Prophet model, developed by Facebook in 2017, is an open-source time-series forecasting algorithm (Taylor et al., 2018). The basic formula for the Prophet model is as follows:

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t \tag{3}$$

Where g(t) represents the trend component, which describes the trend of aperiodic changes in the monthly confirmed SFTS cases time series. s(t) is the periodic component, reflecting the periodic changes in the time series. h(t) represents the impact of holiday events on the time series, and ε_t is the white noise sequence.

2.2.3. LSTM model

LSTM is a special type of recurrent neural network (RNN) that excels in processing and modeling sequence data. Compared to traditional RNN models, LSTM incorporates a gating mechanism, which includes input gates, forget gates, and output gates, to address the long-term dependency problem. By combining neurons, gating units, and cell states, LSTM enables the modeling and memory of long-term dependencies in sequence data.

The LSTM model consists of neurons, where each neuron contains a cell state and three gated units: input gate, forget gate, and output gate. These gated units play a crucial role in controlling the flow and storage of information within the LSTM. The input gate, forget gate, and output gate are responsible for determining which information should be incorporated, forgotten, or outputted by the LSTM. The input gate regulates the inclusion of information from the current time step into the cell state update. If the output of the input gate is close to 1, the corresponding input information is incorporated into the cell state. Conversely, if the output is close to 0, the input information is ignored. Similarly, the forget gate decides which information from the previous cell state should be forgotten. It utilizes a Sigmoid activation function and an element-by-element multiplication operation. If the output of the forget gate is close to 1, the corresponding information is retained. Conversely, if the output is close to 0, the information is forgotten. The cell state is responsible for storing and transmitting information throughout the LSTM. It is updated based on the input gate and forget gate outputs, ensuring relevant information is retained and irrelevant information is discarded. The LSTM structure diagram, as depicted in Fig. 1, visually represents the components and connections within the LSTM model.

The forward propagation process of the LSTM model in the hidden layer unit is as follows.

a. Input gate

$$\mathrm{i_t} = \sigma(\mathrm{W_{xi}x_t} + \mathrm{W_{hi}h_{t-1}} + \mathrm{b_i}) \tag{4}$$

$$\widehat{C}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{5}$$

Where x_t is the input of the current time step, h_{t-1} is the hidden state of the previous time step, W_{xi} and W_{hi} are the weight of the input gate, and b_i is the offset term.

σ, the sigmoid activation function, compresses input to values between 0 and 1, and is commonly used for gate control in LSTMs. When the output is close to 0, it indicates the gate is almost closed, while an output close to 1 indicates the gate is fully open.

Tanh, the hyperbolic tangent activation function, compresses input to values between −1 and 1, and is utilized in LSTMs to generate candidate cell states. These states may be added to the cell state of the next time step based on the activation of the gates.

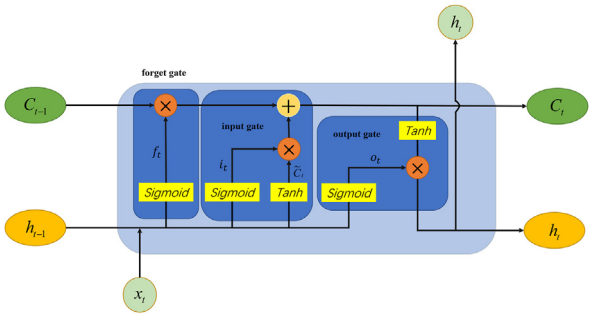

Fig. 1. Brief illustration of LSTM structure.

b. Forget gate

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + b_f\right) \tag{6}$$

Where x_t is the input of the current time step, h_{t-1} is the hidden state of the previous time step, W_{xf} and W_{hf} are the weight of the forget gate, and b_f is the offset term.

c. Output gate

$$O_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{7}$$

$$h_t = O_t \otimes \tanh(C_t) \tag{8}$$

Where x_t is the input of the current time step, h_{t-1} is the hidden state of the previous time step, W_{xo} and W_{ho} are the weight of the output gate, and b_o is the offset term.

d. Cell state

The cell state is responsible for storing and transmitting information. The updating of cell state is determined by the forgetting and input gates. The forgetting gate controls what information from the previous cell state should be forgotten. The output of the forgetting gate is multiplied with the element-by-element of the previous cell state to decide what information to retain. The forgetting step is calculated as follows:

$$C'_{t-1} = f_t \otimes C_{t-1} \tag{9}$$

The input gate controls how new input information affects the updating of the cell state. The output of the input gate is multiplied element-by-element with the new candidate value through an activation function to determine the new cell state. The updated cell state is passed on to the next time step. The calculation formula for the input step is as follows:

$$C_t = C'_{t-1} + i_t \otimes \widehat{C}_t \tag{10}$$

With these two gating mechanisms, LSTM is able to effectively retain useful information and forget irrelevant information.

2.3. Model evaluation

To evaluate the established models, several metrics were used, including the mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE). Each metric has its advantages, with MAE reflecting the actual prediction error, MSE measuring the average sum of squares of the difference between actual and predicted values, and RMSE measuring the deviation from observed values to true values. Lower values of these metrics indicate better prediction performance. The calculation formulas of these four evaluation indicators are as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y}_i| \tag{11}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 \tag{12}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2} \tag{13}$$

$$\mathrm{SMAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2} \tag{14}$$

Where n is the number of observations, y_i is the actual value, $\hat{y}$ is the predicted value.

2.4. Statistical software

Python 3.7 software was used for all statistical and modeling processes, with the "statsmodels" and "fbprophet" packages used to build the SARIMA and Prophet models, respectively. The LSTM model was based on keras, and the level of significance was set at $P < 0.05$.

3. Results

3.1. SFTS cases

From January 2013 to December 2020, a total of 4174 confirmed cases of SFTS were reported in Shandong Province. The monthly confirmed SFTS cases are shown in Fig. 2. The peak of cases began at the beginning of the second quarter and ended at the end of the third quarter, with cases mainly concentrated between May and August.

3.2. Prediction models

3.2.1. SARIMA

To determine the appropriate SARIMA model, the augmented Dickey-Fuller test was conducted to assess the stationarity of the original time series. The test results indicated that the time series was not stationary ($P = 0.313$). After applying a round of differencing, the time series became stationary ($P < 0.001$). The parameters of p and q in the SARIMA model were preliminarily determined based on the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots (Fig. 3). The final SARIMA model was selected based on the minimum Akaike information criterion (AIC) value (AIC = 534.757), resulting in SARIMA $(1,1,1)$, $(1,1,1)_{12}$. The residual autocorrelation test (Ljung-Box test) indicated that the residuals did not deviate significantly from a white noise sequence ($P = 0.89$), confirming the adequacy of the model.

3.2.2. Prophet

The seasonality_mode parameter in the Prophet model determines the modeling approach for the seasonal component and can be set to additive or multiplicative. Considering the increasing magnitude of the seasonal pattern with the growth of data, which aligns with the characteristics of a multiplicative model, we selected multiplicative as the parameter for seasonality_mode. Table 1 presents all the parameters used in the constructed Prophet model.

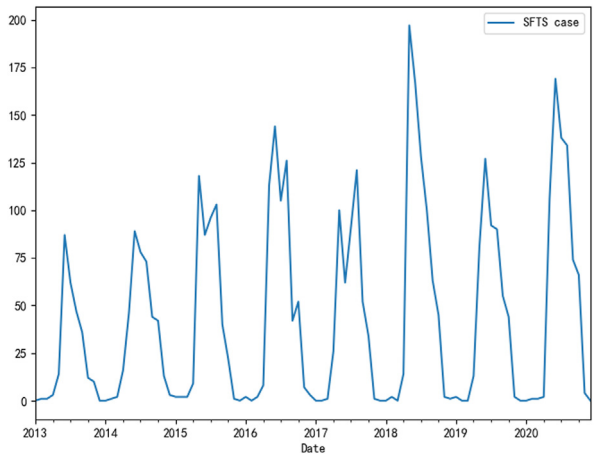

Fig. 2. Trends of the actual number of SFTS cases from January 2013 to December 2020.

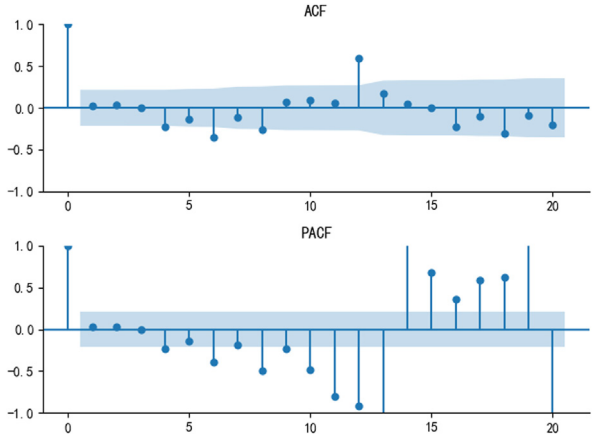

Fig. 3. Autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of the SARIMA (1,1,1), $(1,1,1)_{12}$ model.

Table 1
Parameter values of the constructed Prophet and LSTM models.

Models	Parameters	Values
Prophet	growth	linear
	yearly_seasonality	True
	seasonality_mode	multiplicative
	seasonality_prior_scale	12
	changepoint_prior_scale	0.02
LSTM	Number of neurons	23
	layers	1
	Activation	relu
	Recurrent activation	sigmoid
	Dropout	0
	Loss	mse
	Optimizer	Adam
	Batch size	1
	Learning_rate	0.003
	Epochs	1000

3.2.3. LSTM

To prepare the input time series for the LSTM model, it was first transformed into the required input-output format for the supervised learning model. The lengths of both input and output sequences were then defined and normalized for optimal model training.

The Bayesian optimization method was utilized to automatically search for the optimal learning rate and number of neurons. An evaluation function was defined to maximize the R-squared value, while the remaining parameters were adjusted manually. To prevent overfitting and improve the model's ability to generalize to unseen data, the early stopping technique was introduced. Halting training when the model's performance on the validation set started to deteriorate helped prevent overfitting. The "patient" parameter was set to 100, meaning that if the validation loss did not improve for 100 consecutive epochs, training would stop, and the model with the best weights based on the validation loss would be saved. Detailed parameters of the LSTM model are outlined in Table 1.

If the forecast yielded a negative value, it was adjusted to 0 to align with the actual scenario using the np.clip function from the NumPy library, which performed a clipping operation on a numerical array, restricting its elements to a specified range.

As shown in Fig. 4, the training loss and validation loss displayed a decreasing trend in loss values as the number of training rounds increased. In the initial stage of training, the loss rapidly decreased, indicating that the model was learning quickly. As the number of training rounds increased, the rate of loss decrease slowed down, and the curve tended to flatten, indicating that the model was gradually converging. The small difference between training loss and validation loss indicated that the model had not experienced overfitting or underfitting. Overall, the training process of this LSTM model appeared to be healthy and the model had good generalization capabilities on both the training and validation datasets. Based on this model, the number of SFTS cases in the subsequent 5 years in this area were generated (Supplementary file 1).

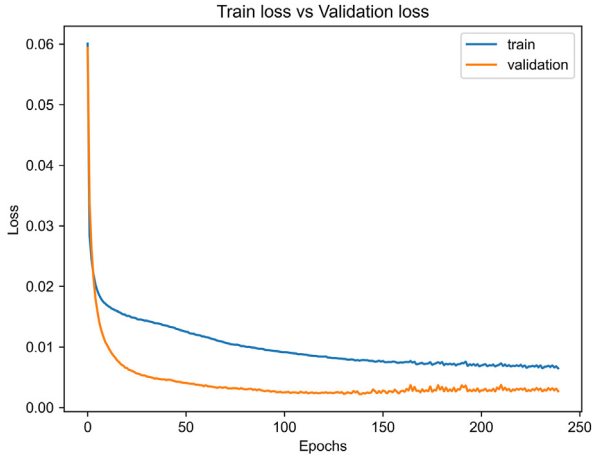

Fig. 4. Loss function for the LSTM model.

3.3. Model comparison

The predicted and actual SFTS cases in 2020 are presented in Table 2 and Fig. 5. All three constructed models performed well in predicting SFTS cases. The SARIMA model generated the closest number of cases to the actual cases in May and October, while the Prophet model performed best in January, June, and November. The LSTM model excelled in February, March, July, August, September, and December. However, the predicted number of cases in April by all the models was considerably higher than the actual number. To further compare and evaluate these models, various error analysis methods, including MAE, MAPE, and RMSE, were employed. As shown in Table 3, the LSTM model demonstrated the best fitting and prediction performance, with the lowest values for MAE, MAPE, and RMSE. It was followed by the Prophet and SARIMA models, respectively.

4. Discussion

The region of Shandong Province, where this study was conducted, has been identified as one of the areas with a high incidence of SFTS (Huang et al., 2021; Liu et al., 2015). Previous studies have analyzed the risk factors associated with SFTS incidence and developed multivariate models in this region (Hou et al., 2023; Jiang et al., 2022; Wang et al., 2022). However, to our knowledge, this is the first study to focus on building univariate models for predicting SFTS incidence, which is simpler and more practical.

All three constructed models performed well in predicting SFTS cases and showed similar trends to the actual cases. However, they did not accurately predict the cases in April. This could be due to the sudden increase in actual cases during this month, and the models may have difficulty predicting volatile data.

The ARIMA and SARIMA models have the advantage of being relatively simple linear models and can capture the dynamic relationships between historical and predicted data. However, these models have some limitations. They require stable data and struggle to capture nonlinear relationships in the data. Among the three models, SARIMA performed the worst, especially when predicting cases during months of high prevalence (June to August).

In contrast to ARIMA and SARIMA, the Prophet model does not require the time series data to be stationary before modeling and offers more adjustable parameters, making it more flexible. Since its introduction, the Prophet model has been widely used in medical research and other areas, such as COVID-19 (Battineni et al., 2020; Khayyat et al., 2021; Satrio et al., 2021), hand, foot and mouth disease (Xie et al.), air pollution (Shen et al., 2020), road traffic injuries (Feng et al., 2022), etc., and has shown good prediction performance. In this study, the Prophet model performed better when the actual number of cases was higher or lower (e.g., in January, June, and November).

Traditional RNNs are prone to the problems of vanishing or exploding gradients when dealing with long sequences. However, LSTM can effectively alleviate these issues by introducing a gating mechanism. LSTM has memory units that store and read information at different time steps, making it suitable for handling long-term dependencies and processing long sequences. In this study, LSTM performed the best with the lowest error values, although it did not perform as well as Prophet in predicting cases in July, when the number of actual cases was the highest.

The incidence data after 2021 were not public for now, and the number of SFTS cases in the subsequent 5 years (2021–2025) in this area were predicted using the established LSTM model, which can be used as a reference for preparing medical supplies in advance, formulating targeted prevention and control measures, and providing early warnings of SFTS prevalence.

Table 2
Predicted monthly number of SFTS cases in 2020 by three models.

Models	SARIMA	Prophet	LSTM	Actual
January	7.7	0.8	0.6	0
February	7.4	3.5	3.5	1
March	7.6	2.6	2.6	1
April	20.8	16.3	15.1	2
May	117.3	136.8	125.7	106
June	113.3	154.2	166.4	169
July	123.0	127.0	117.6	138
August	120.6	146.3	147.0	134
September	68.6	53.1	69.2	74
October	54.6	49.8	71.8	66
November	20.0	5.0	4.4	4
December	8.0	1.9	0.0	0

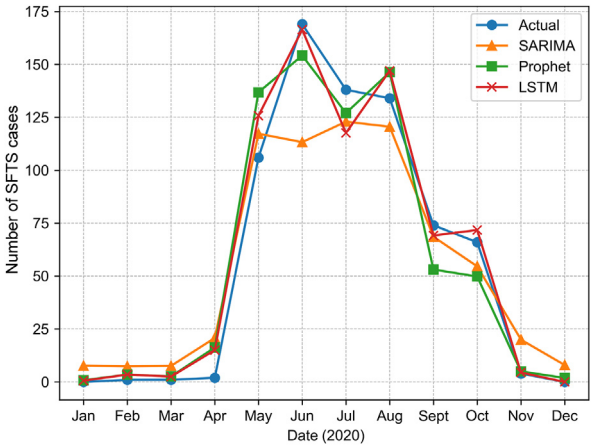

Fig. 5. Comparison of the actual SFTS cases with the predicted cases by the three models from January 2020 to December 2020.

Table 3
Comparison of three models using mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE).

Models	SARIMA	Prophet	LSTM
MAE	14.64	10.67	7.04
MSE	384.30	196.27	101.41
RMSE	19.60	14.01	10.07

There are some limitations to this study. There may be discrepancies between the forecasted data and the actual data in certain months. Some studies have shown that combining linear and nonlinear models may yield better predictive performance than using a single model, such as SARIMA-Prophet(Luo et al., 2022), Prophet-LSTM (Na et al., 2019), SARIMA-NARX (Wang et al., 2019), etc. Therefore, further research is needed to enhance the predictive ability. In addition, this study focused on the cases in Shandong province, and the universality of this model in other epidemic areas still needs further evaluation.

5. Conclusions

In conclusion, this study developed three univariate prediction models based on historical monthly SFTS cases. Among these models, the LSTM model performed the best in predicting the monthly confirmed SFTS cases in Shandong Province, China. This simple and practical model can provide valuable information and data for assessing the potential risk of SFTS in advance, thereby benefiting early warning systems and the formulation of prevention and control measures for SFTS.

Ethics approval and consent to participate

The use of human data was approved by the Ethics Committee of Huadong Research Institute for Medicine and Biotechniques.

Consent for publication

Not applicable.

Availability of data and materials

The datasets during and/or analyzed during the current study available from the corresponding author on reasonable request.

Funding

This study was funded by Medical Science and Technology Projects, China (JK2023GK002, JK2023GK003, and JK2023GK004).

CRediT authorship contribution statement

Zixu Wang: Writing – original draft, Software, Methodology, Data curation, Conceptualization. **Wenyi Zhang:** Resources, Data curation, Conceptualization. **Ting Wu:** Methodology. **Nianhong Lu:** Formal analysis. **Junyu He:** Writing – review & editing, Resources. **Junhu Wang:** Formal analysis. **Jixian Rao:** Formal analysis. **Yuan Gu:** Formal analysis. **Xianxian Cheng:** Software. **Yuexi Li:** Funding acquisition. **Yong Qi:** Writing – review & editing, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.idm.2024.01.003.

References

Battineni, G., Chintalapudi, N., & Amenta, F. (2020). Forecasting of COVID-19 epidemic size in four high hitting nations (USA, Brazil, India and Russia) by Fb-Prophet machine learning model. *Applied Computing and Informatics, 6*(1), 1–10.

Cho, G., Lee, S., & Lee, H. (2021). Estimating severe fever with thrombocytopenia syndrome transmission using machine learning methods in South Korea. *Scientific Reports, 11*(1), Article 21831. https://doi.org/10.1038/s41598-021-01361-9

Deng, B., Rui, J., Liang, S. Y., Li, Z. F., Li, K., Lin, S., Luo, L., Xu, J., Liu, W., Huang, J., Wei, H., Yang, T., Liu, C., Li, Z., Li, P., Zhao, Z., Wang, Y., Yang, M., Zhu, Y., … Chen, T. (2022). Meteorological factors and tick density affect the dynamics of SFTS in jiangsu province, China. *PLoS Negl Trop Dis, 16*(5), Article e0010432. https://doi.org/10.1371/journal.pntd.0010432

Dharmarajan, G., Li, R., Chanda, E., Dean, K. R., Dirzo, R., Jakobsen, K. S., Khan, I., Leirs, H., Shi, Z.-L., Wolfe, N. D., Yang, R., & Stenseth, N. C. (2022). The Animal Origin of Major human infectious diseases: What can Past Ep idemics Teach Us about preventing the next Pandemic? *Zoonoses, 2*(1). https://doi.org/10.15212/zoonoses-2021-0028

Ding, F., Guan, X. H., Kang, K., Ding, S. J., Huang, L. Y., Xing, X. S., Sha, S., Liu, L., Wang, X. J., Zhang, X. M., You, A. G., Du, Y. H., Zhou, H., Vong, S., Zhang, X. D., Feng, Z. J., Yang, W. Z., Li, Q., & Yin, W. W. (2014). Risk factors for bunyavirus-associated severe Fever with thrombocytopenia syndrome, China. *PLoS Negl Trop Dis, 8*(10), e3267. https://doi.org/10.1371/journal.pntd.0003267

Feng, T., Zheng, Z., Xu, J., Liu, M., Li, M., Jia, H., & Yu, X. (2022). The comparative analysis of SARIMA, Facebook Prophet, and LSTM for road traffic injury prediction in Northeast China. *Frontiers in Public Health, 10*, Article 946563. https://doi.org/10.3389/fpubh.2022.946563

Hou, S., Zhang, N., Liu, J., Li, H., Liu, X., & Liu, T. (2023). Epidemiological characteristics and risk factors of severe fever with thrombocytopenia syndrome in Yantai City, Shandong province. *Open Forum Infectious Diseases, 10*(4), ofad141. https://doi.org/10.1093/ofid/ofad141

Huang, X., Li, J., Li, A., Wang, S., & Li, D. (2021). Epidemiological characteristics of severe fever with thrombocytopenia syndrome from 2010 to 2019 in mainland China. *Int J Environ Res Public Health, 18*(6). https://doi.org/10.3390/ijerph18063092

Jiang, X., Wang, Y., Zhang, X., Pang, B., Yao, M., Tian, X., & Sang, S. (2022). Factors associated with severe fever with thrombocytopenia syndrome in Endemic areas of China. *Frontiers in Public Health, 10*, Article 844220. https://doi.org/10.3389/fpubh.2022.844220

Kato, H., Yamagishi, T., Shimada, T., Matsui, T., Shimojima, M., Saijo, M., & Oishi, K. (2016). Epidemiological and clinical features of severe fever with thrombocytopenia syndrome in Japan, 2013-2014. *PLoS One, 11*(10), Article e0165207. https://doi.org/10.1371/journal.pone.0165207

Khayyat, M., Laabidi, K., Almalki, N., & Zahrani, M. A. (2021). *Time series Facebook Prophet model and Python for COVID-19 Outbreak prediction*. Cmc -Tech Science Press-.

Li, H., Lu, Q. B., Xing, B., Zhang, S. F., Liu, K., Du, J., Li, X. K., Cui, N., Yang, Z. D., Wang, L. Y., Hu, J. G., Cao, W. C., & Liu, W. (2018). Epidemiological and clinical features of laboratory-diagnosed severe fever with thrombocytopenia syndrome in China, 2011-17: A prospective observational study. *The Lancet Infectious Diseases, 18*(10), 1127–1137. https://doi.org/10.1016/s1473-3099(18)30293-7

Liu, K., Zhou, H., Sun, R. X., Yao, H. W., Li, Y., Wang, L. P., Di, M., Li, X. L., Yang, Y., Gray, G. C., Cui, N., Yin, W. W., Fang, L. Q., Yu, H. J., & Cao, W. C. (2015). A national assessment of the epidemiology of severe fever with thrombocytopenia syndrome, China. *Scientific Reports, 5*, 9679. https://doi.org/10.1038/srep09679

Luo, Z., Jia, X., Bao, J., Song, Z., Zhu, H., Liu, M., Yang, Y., & Shi, X. (2022). A combined model of SARIMA and Prophet models in forecasting AIDS incidence in Henan province, China. *Int J Environ Res Public Health, 19*(10). https://doi.org/10.3390/ijerph19105910

McMullan, L. K., Folk, S. M., Kelly, A. J., MacNeil, A., Goldsmith, C. S., Metcalfe, M. G., Batten, B. C., Albariño, C. G., Zaki, S. R., Rollin, P. E., Nicholson, W. L., & Nichol, S. T. (2012). A new phlebovirus associated with severe febrile illness in Missouri. *New England Journal of Medicine, 367*(9), 834–841. https://doi.org/10.1056/NEJMoa1203378

Na, G. E., Lian-Ying, S., Xiao-Da, S., Ping, Z., & University, B. U. (2019). Research on Sales forecast of Prophet-LSTM Combination model. *Computer Science*.

Satrio, C. B. A., Darmawan, W., Nadia, B. U., & Hanafiah, N. (2021). Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. *Procedia Computer Science, 179*(12), 524–532.

Shen, J., Valagolam, D., & Mccalla, S. (2020). Prophet forecasting model: A machine learning approach to predict the concentration of air pollutants (PM2.5, PM10, O3, NO2, SO2, CO) in Seoul, South Korea. *PeerJ, 8*(3), Article e9961.

Shin, J., Kwon, D., Youn, S. K., & Park, J. H. (2015). Characteristics and factors associated with Death among patients Hospitalized for severe fever with thrombocytopenia syndrome, South Korea, 2013. *Emerging Infectious Diseases, 21*(10), 1704–1710. https://doi.org/10.3201/eid2110.141928

Sun, J. M., Lu, L., Liu, K. K., Yang, J., Wu, H. X., & Liu, Q. Y. (2018). Forecast of severe fever with thrombocytopenia syndrome incidence with meteorological factors. *Sci Total Environ, 626*, 1188–1192. https://doi.org/10.1016/j.scitotenv.2018.01.196

Sun, J. M., Wu, H. X., Lu, L., Liu, Y., Mao, Z. Y., Ren, J. P., Yao, W. W., Qu, H. H., & Liu, Q. Y. (2021). Factors associated with spatial distribution of severe fever with thrombocytopenia syndrome. *Sci Total Environ, 750*, Article 141522. https://doi.org/10.1016/j.scitotenv.2020.141522

Wang, Y., Pang, B., Ma, W., Kou, Z., & Wen, H. (2022). Analysis of the spatial-temporal components driving transmission of the severe fever with thrombocytopenia syndrome in Shandong Province, China, 2016-2018. *Transboundary and emerging diseases, 69*(6), 3761–3770. https://doi.org/10.1111/tbed.14745

Wang, Y., Xu, C., Wang, Z., & Yuan, J. (2019). Seasonality and trend prediction of scarlet fever incidence in mainland China from 2004 to 2018 using a hybrid SARIMA-NARX model. *PeerJ, 7*, Article e6165. https://doi.org/10.7717/peerj.6165

Wu, H., Wu, C., Lu, Q., Ding, Z., Xue, M., & Lin, J. (2020). Spatial-temporal characteristics of severe fever with thrombocytopenia syndrome and the relationship with meteorological factors from 2011 to 2018 in Zhejiang Province, China. *PLoS Negl Trop Dis, 14*(4), Article e0008186. https://doi.org/10.1371/journal.pntd.0008186

Xie, C., Wen, H., Yang, W., Cai, J., Zhang, P., Wu, R., Li, M., Huang, S., Trend analysis and forecast of daily reported incidence of hand, foot and mouth disease in Hubei, China by Prophet model. Scientific Reports.

Yu, X. J., Liang, M. F., Zhang, S. Y., Liu, Y., Li, J. D., Sun, Y. L., Zhang, L., Zhang, Q. F., Popov, V. L., Li, C., Qu, J., Li, Q., Zhang, Y. P., Hai, R., Wu, W., Wang, Q., Zhan, F. X., Wang, X. J., Kan, B., … Li, D. X. (2011). Fever with thrombocytopenia associated with a novel bunyavirus in China. *New England Journal of Medicine, 364*(16), 1523–1532. https://doi.org/10.1056/NEJMoa1010095