Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 4, No. 2, 2025

The Prediction of Influenza Using the Hybrid ARIMA-LSTM Model

Yi Zou*

University of Putra Malaysia, Serdang 43400, Malaysia
* **Corresponding author**: Yi Zou

Abstract: Time series analysis plays an important role in many fields. Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) methods are the most common tools to forecast sequential and time series data. However, both of them have obvious defects. The ARIMA model cannot capture non-linear information in sequences. Although the LSTM network is good at learning dynamic variations, it is prone to overfitting and requires a mass of long-term data. In this study, a hybrid technology blending the ARIMA and the LSTM algorithms was utilized to forecast the number of cases of influenza in China. This hybrid method leveraged the advantages of the ARIMA model and the LSTM network. Firstly, the ARIMA model was used to analyze the linear relationship within the time series. Then, the residuals of the ARIMA were taken as the input values to train the LSTM networks. A new hybrid ARIMA-LSTM model obtained, combining a SARIMA (0,1,0) (1,0,0)52 model and a LSTM model with 50 epochs and 32 batch size. This model successfully addressed the previously mentioned issues and enhanced the precision of the predictions. It managed to reduce 4.6% of RMSE, 8.9% of MSE, and 13.9% of MAE. In addition, this new algorithm was found that it didn't have a high requirement like the individual LSTM model. Since there were not very many observations in the dataset, the performance of the individual LSTM model was not good. However, the integrated model improved this problem and obtained a more precise prediction. Even though the hybrid model had a better performance on prediction, it still has the risk of overfitting the data. The future work will be to improve the hybrid model to decrease this risk by adding more variables and modifying the structure of the LSTM model. Meanwhile, applying this method to another field further proves its feasibility and provides more effective prediction models.

Keywords: Time Series, ARIMA, Artificial Neural Network, LSTM, Influenza, Prediction

1. Introduction

Time series analysis is an important area of predictive analytics, involving the collection of previous observations of a variable in time sequence to analyze data characteristics and build models for forecasting future values. It is a highly effective analytical method when little is known about the process that generated the object of study or when relevant data are limited. Time series analysis is utilized in many domains including economic forecasting, stock market analysis, weather forecasting, earthquake prediction, control engineering, astronomy, heart rate monitoring, electroencephalography, etc.

The accuracy of forecasting models is important for the following reasons. First, forecasts can be used to guide both short-term and long-term decision-making. Second, forecasts help deal with uncertainty in the data. If the model is inaccurate and leads to incorrect decisions, it can result in manpower, material, and economic losses.

There are many ways to forecast time series data. In this study, the Autoregressive Integrated Moving Average (ARIMA) model and the Long Short-Term Memory (LSTM) recurrent neural network were applied.

In time series analysis, the Autoregressive Integrated Moving Average (ARIMA) model is one of the most widely used and significant models. Yule, Slutsky, Walker, and Yaglom formulated the concept of Autoregressive (AR) and Moving Average (MA) models. The ARIMA model was developed from the Autoregressive Moving Average (ARMA) model. A stationary time series has properties independent of the time at which the series is observed; in contrast, a time series with trend or seasonality is not stationary, as the trend and seasonality affect the values of the time series at different times. When dealing with practical problems, it is widespread to encounter non-stationary original data. The ARIMA model incorporates a differencing step to smooth out the non-stationary series to build the model. In 1975, the general transfer function model of the ARIMA procedure was proposed by Box and Tiao. These models provide accurate predictions for linear data and can be adjusted and optimized according to the characteristics of the series, while their predictive capability is limited for non-linear data.

Unlike the ARIMA model, the Artificial Neural Network (ANN) is proficient at handling non-linear time series data. ANNs are adaptive, data-driven methods with few a priori assumptions about the models for the problems under study. And ANNs can generalize. After learning from the data (samples) presented to it, artificial neural networks can often correctly infer unseen parts of the population, even if the sample data contains noisy information. Besides, ANNs are universal functional approximators, capable of approximating any continuous function to any desired accuracy. Lapedes and Farber (1987) were the first to attempt modeling non-linear time series using artificial neural networks. Chakraborty, Mehrotra, Mohan, and Ranka (1992) conducted an empirical study on multivariate time series forecasting with artificial neural networks. Poli and Jones proposed a stochastic neural network model based on the Kalman filter for non-linear time series prediction in 1994. Berardi and Zhang (2003) studied the bias and variance issues in the context of time series forecasting. Over the past two decades, ANN has been continuously developed and widely applied in time series prediction.

Recurrent Neural Network (RNN) is a specific artificial neural network used to process time series data and sequential information. Unlike traditional neural networks, RNNs have an element of memory and can use historical information to evaluate current inputs. Due to these properties, RNNs are very effective in modeling time dependencies and sequential relationships. The purpose of RNN is to predict the next step in a sequence of observations based on the previous steps observed in the sequence. RNN utilizes successive observations and learns from the earlier stages to predict future trends. In the early stages, data need to be remembered to guess the next step. In RNN, the hidden layer acts as an internal memory to store information collected in the early stages of processing sequential data.

Long Short-Term Memory (LSTM) neural network is a type of Recurrent Neural Network (RNN) uses besides standard units, special units. The LSTM model is designed to address the vanishing gradient problem. It is capable of recalling previous long-term time-series data and automatically controlling the retention of relevant features and discarding irrelevant features in data.

Influenza is an acute respiratory infectious disease caused by influenza virus and the first epidemic to be monitored worldwide. It is common in all parts of the world. Influenza spreads easily between people when they cough or sneeze. Most people recover from fever and other symptoms within a week without requiring medical attention. However, influenza can cause severe illness or death, especially in people at high risk.

Influenza can worsen symptoms of other chronic diseases. In severe cases, it can lead to pneumonia and sepsis. Epidemics can result in high levels of worker/school absenteeism and productivity losses. Clinics and hospitals can be overwhelmed during peak illness periods.

2. The shortages of Existing flu forecasting methods

Autoregressive Integrated Moving Average (ARIMA) model, is a time series autoregressive technique that generates short-term predictions by analyzing historical time series data. The ARIMA model has been used to predict outbreaks of many diseases in the past such as Hepatitis B, Tuberculosis, AIDS, COVID-19, as well as Influenza. ARIMA model has some advantages: First, it is suitable for predicting various types of time series data, as long as the data exhibits a certain degree of stability and autocorrelation; Two, it can capture different types of time series characteristics like seasonality, trend and stationarity by adjusting parameters. Three, it is a relatively simple model because there are only involved 3 parameters in this model. On the other hand, the ARIMA model is linear, which is not able to handle non-linear data effectively. Meanwhile, its prediction accuracy depends on historical data: The ARIMA model assumes that the characteristics of future data are similar to historical data. When the data structure changes, the accuracy of the model prediction will decrease.

Recurrent Neural Networks (RNNs) have been widely used in the study of sequential data such as text, audio, and video. However, when the input gap is large, RNNs composed of sigma units or tanh units cannot learn the relevant information of the input data. By introducing gate functions in the unit structure, Long Short-Term Memory (LSTM) neural networks can handle long-term dependency problems well. They can learn the non-linear and non-stationary nature of a time series, making them capable of predicting future trends and reducing forecasting errors after training. However, The LSTM model requires a significant amount of data for training to understand complex patterns, making it computationally expensive for training and evaluation.

For complex data with both linear and non-linear trends, the model fitting effect will be limited by establishing individual ARIMA models or individual LSTM models separately. In recent years, some studies have proposed to build the ARIMA-LSTM joint model improving the accuracy of predictions to process this kind of problem.

ARIMA models and LSTM models are also widely used in disease prediction, especially epidemic prediction, because epidemics have the characteristic of time spread. Therefore, influenza is an important application for them.

Influenza is the first infectious respiratory disease to be monitored globally. It is a viral disease that spreads rapidly across regions through human-to-human transmission and has an extremely high mortality rate in specific populations. According to WHO statistics, annual influenza epidemics cause 5%-15% of the world's population to suffer from respiratory diseases, 3 million to 5 million severe cases, and 250,000 to 500,000 deaths. The United States suffers economic losses of $71 billion to $167 billion each year due to influenza.

However, most of studies of flu applied separately one of these two algorithms. Even in other medical fields, there are few studies combining them. This situation limits the ability of the prediction model to capture data information, thus reducing the accuracy of predictions.

In the stock or energy fields, the ARIMA-LSTM method has been developed few years. Many researchers continuously propose and determine the feasibility of it. And they indicate this mixing technique has more power to detect and forecast the sequential and time series data.

3. Methodology

3.1. Data Preprocessing

3.1.1. Data Collection

In this study, weekly data on received Influenza specimens' cases in China was retrieved from the World Health Organization website. The dataset spans from July 2009 to February 2024.

3.1.2. Missing Value Imputation

In data analysis, it is quite common that raw data are incomplete due to various reasons. Since the dataset covers more than a decade and China has over a billion people, the difficulties of data collection and data storage are inevitably increasing. After checking, 14 missing values were detected, and the proportion of total data is very small. When there are only a few missing data in large-size datasets, it is a conventional approach to delete them. Therefore, in this case, these 14 missing values were removed.

3.1.3. Stationarity Test

In time series analysis, data stationarity is very important, if not it is quite possible to appear spurious regression (Rizwan, 2011). The Augmented Dickey-Fuller (ADF) test is a powerful way to determine whether a time series is stationary, i.e., it has a unit root.

The assumption is as:

H0: the series has a unit root ($\gamma = 0$).

H1: the series does not have a unit root ($\gamma < 0$).

ADF test is basic on the model:

$$\Delta y_t = \alpha + \beta_t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} \dots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \quad (1)$$

where α is a constant; β is the coefficient on a time trend; p is the lag order of the autoregressive process; Δ is the different operator; γ is the coefficient defining the unit root; δ is the coefficient of the difference of series; ε is the error term.

The test statistic for the ADF has the following form:

$$DF_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \quad (2)$$

where SE is the standard error of the least squares estimate of the γ coefficient. The null hypothesis is rejected if $DF\tau$ less (more negative) than the critical value, in other words, the time series is stationary.

3.2. Autoregressive Model

It is rare for a time series to be completely random. Instead, the observation within a specific interval can be expected to rely on the preceding observations. Consequently, the autoregressive model (AR model), a form of linear regression, is applied to lagged series derived from the original time series. The term 'autoregression' indicates a regression where a variable is regressed against its past values. In this model, the output is the future data point and can be expressed as a linear combination of past p data points, where p is the lag.

Thus, an autoregressive model of order p is given by:

$$yt = \varphi 1yt\text{-}1 + \varphi 2yt\text{-}2 + \dots + \varphi pyt\text{-}p + \varepsilon t \quad (3)$$

where {yt} is a time series, φ are the autoregressive coefficients, p is the autoregressive order, ε is a white noise with zero mean, constant variance and no autocorrelation.

3.3. Moving Average Model

The relationship between past white noise error terms and present values of the time series is frequently observed due to the occurrence of unforeseen events. Accordingly, the Moving Average model (MA model) differs from the AR model in that it relies on the past white noise error terms of the time series instead of past values of the variable predicted.

The MA(q) model is described as:

$$yt = \varepsilon t + \theta 1\varepsilon t\text{-}1 + \theta 2\varepsilon t\text{-}2 + \dots + \theta q\varepsilon t\text{-}q \quad (4)$$

where {yt} represents a time series, θ denotes the moving average coefficients, q denotes the order of the moving average, and ε represents a white noise with zero mean, constant variance, and no autocorrelation.

3.4. Autoregressive Integrated Moving Average Model

The Autoregressive Integrated Moving Average (ARIMA) method predicts the current value in a time series by integrating its past values and random errors (also known as shocks or innovations). In general, nonseasonal times series can be represented as ARIMA (p, d, q):

$$yt = \varphi 1yt\text{-}1 + \varphi 2yt\text{-}2 + \dots + \varphi pyt\text{-}p + \varepsilon t + \theta 1\varepsilon t\text{-}1 + \theta 2\varepsilon t\text{-}2 + \dots + \theta q\varepsilon t\text{-}q \quad (5)$$

where {yt} represents a time series; φ and θ respectively denote the autoregressive coefficients and the moving average coefficients; p denotes the order of the autoregressive part; d denotes the order of differencing level; q and the order of the moving average part (Nevil and Rajendra, 2021); ε represents a white noise with zero mean, constant variance, and no autocorrelation.

This approach was originally generalized by Box and Jenkins (1970), called the Box-Jenkins procedure. The ARIMA model modeling includes the following steps:

1) Check the stationarity of the time series.

If the data is non-stationary, differencing or detrend should be done before modeling, i.e., d is determined.

2) Identify p and q.

Preliminary determine model by ACF plot and PACF plot.

a) The Autocorrelation (ACF) plot:

Plotting the correlation between the observation and its previous values over different lag lengths in time. This plot is used to determine the order of the MA process.

b) The Partial Autocorrelation (PACF) plot:

Plotting the partial correlation of a stationary time series with its own lagged values over a shorter period. This plot is used to confirm the order of the AR process.

Refer to Cryer and Chan (2008), Table 1 shows how to decide the order of the model:

Table 1 Principle of the Determine of Orders

	AR (p)	MA (q)	ARMA(p,q)
ACF Plot	Tails off	Cuts off after lag q	Tails off
PACF Plot	Cuts off after lag p	Tails off	Tails off

3) Then comparison of AIC and BIC.

The Akaike information criterion (AIC) and the Bayes information criterion (BIC) are criteria to select the suitable model from candidate models. Essentially, they settle down to avoid overfitting or underfitting. When selecting a model, the one with the lower AIC or BIC will be preferred.

The formula of AIC is the following:

$$AIC = 2K - 2\ln(\hat{L}) \quad (6)$$

The formula of BIC is the following:

$$BIC = -2\log L + K \log N \quad (7)$$

where L is the likelihood, K is the number of parameters, and N is the number of data (Yang and Yang, 2014).

4) Parameter estimation

Maximum likelihood estimation (MLE) is a popular method to estimate the parameters of models. It works by maximizing the probability of the sample data to get the parameter values.

3.5. Seasonal Autoregressive Integrated Moving Average Model

Seasonal Autoregressive Integrated Moving Average (SARIMA) is an extension of the ARIMA model, which combines the seasonal, autoregressive, intergraded, and moving average components. To time series data with seasonal variations, short-run non-seasonal components likely contribute to the model. Therefore, we need to estimate the seasonal ARIMA model, which incorporates both non-seasonal and seasonal factors in a multiplicative model.

The SARIMA model is represented as ARIMA (p, d, q) (P, D, Q) s, where P is the order of the seasonal AR component; D is the order of seasonal differences; Q is the order of the seasonal MA component; s is the period length (Nevil and Rajendra, 2021).

The SARIMA can be written as:

$$\phi P(B)\Phi P(Bs)(1-B)\mathrm{d}(1-Bs)\mathrm{DXt}=\theta \mathrm{q}(B)\Theta \mathrm{Q}(Bs)\varepsilon t \quad (8)$$

where

$$\mathrm{BXt} = \mathrm{Xt\text{-}1} \quad (9)$$

$$\phi p(B) = (1-\phi 1B-\ldots-\phi pBp) \quad (10)$$

$$\theta q(B) = (1-\theta 1B-\ldots-\theta qBq) \quad (11)$$

$$\Phi P(Bs) = (1-\Phi 1Bs-\ldots-\Phi PBsP) \quad (12)$$

$$\Theta Q(Bs) = (1-\Theta 1Bs-\ldots-\Theta QBsQ) \quad (13)$$

3.6. Artificial Neural Network

Artificial neural network (ANN) is a vital deep learning method, mimicking biological nervous systems to transfer, process, and recognize information for decision-making ANNs are regarded as non-linear function approximators due to map non-linear correspondence. This algorithm can learn and train neural networks to identify hidden patterns and underlying dynamics (Selvin, et al. 2017).

Figure 1 shows the structure of ANN. A neural network consists of nodes connected through at least three layers:

1) Input layer:

This layer receives the original input data. Each neuron or node denotes a feature or dimension of the input data. The number of nodes in this layer determines the dimensionality of the input data.

2) Hidden layers:

This layer is in the middle of the input layer and the output layer, there could be more than one layer. In these layers, the neurons transfer the values to the output layer by the activation function with a weighted sum of the inputs. The purpose of the activation function is to introduce non-linearity into the neural network thereby increasing the ability to learn complicated patterns. The usual activation functions include sigmoid, tangent hyperbolic (tanh), Rectified Linear Unit (ReLU), etc.

3) Output layer:

This layer outputs the final prediction processing from the hidden layers.

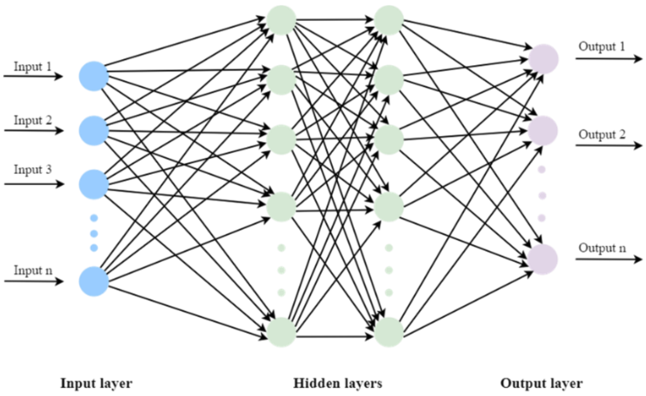

Figure 1 Structure of ANN

3.7. Recurrent Neural Network

Recurrent Neural Network (RNN), a specific type of ANN, is designed for sequential data. One of the simplest and most basic ANN is the Feedforward Neural Network where data propagate in one direction from input layers to output layers. RNN adds memory or feedback loops into the original ANN system on the hidden layer so cell states (cell memory) are passed from the previous timestep to the next. Accordingly, RNNs are appropriate for time series data.

In general, the Backpropagation Through Time algorithm is used to train the RNN system with the gradient optimization descent to learn the weights. In this algorithm, the key is the error backward. First, the unfolded neural network is constructed, and the calculated output values are obtained with random incipient weights. Sequentially, there are differences between the calculated output values and the original output values. Then, updating the weights and bias makes use of reducing the error until the minimum error overall timesteps (Ralf and Eric, 2019).

Figure 2 shows the transformation of RNN and Figure 3 is the detailed structure of RNN.

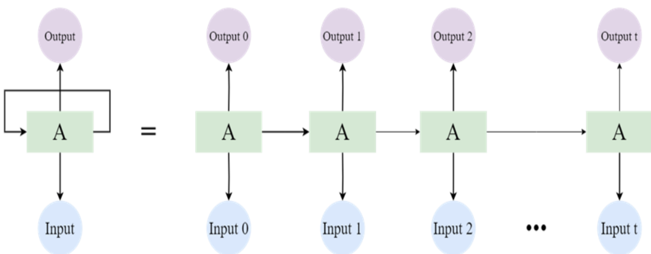

Figure 2 Transformation of RNN

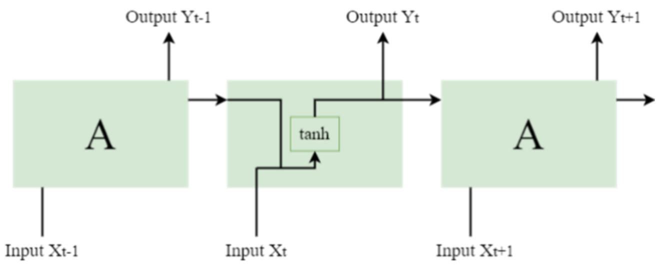

Figure 3 Structure of RNN

3.8. Long Short-Term Memory Network

LSTM (Long Short-Term Memory) neural network is an iterative structure in the hidden layer of the recurrent neural network that could capture the long-term variation and dependency in sequence and time series leading to solving the vanishing gradient and exploding gradient. And LSTM is capable of memorizing and transferring previous information. LSTM introduces the memory cell that consists of three components: input gate, forget gate, and output gate. Figure 4 shows the structure of a single memory cell. The cell state carries and transfers the information from one memory cell to another one. The gates control which information should be kept and which information should be taken out. LSTM does not like RNN which only has one layer of tanh, it has four interacting layers. Hence, the LSTM neural network overcoming RNN's defects plays an outstanding role in learning and forecasting sequence and time series data.

The input gate decides what information should be stored by the memory cell, as in:

$$i_t = \sigma(W_i X_t + U_i h_{t-1} + b_i) \tag{14}$$

where Xt is the input vector; it is the input gate vector; W, U, and b are parameter matrices.

The forget gate decides what information needs to be removed as:

$$f_t = \sigma\left(W_f X_t + U_f h_{t-1} + b_f\right) \tag{15}$$

where ft is the forget gate vector.

The output gate decides what information is outputted from the memory cell as

$$O_t = \sigma(W_o X_t + U_o h_{t-1} + b_o) \tag{16}$$

where Ot is the output gate vector.

The new information will be updated into the new cell state by the current input cell state as

$$\overline{C}_t = \sigma(W_c X_t + U_c h_{t-1} + b_c) \tag{17}$$

where $\overline{C}_t$ is the current input cell state; ht-1 is the hidden state vector at time t-1.

$$C_t = f_t C_{t-1} + i_t \overline{C_t} \tag{18}$$

where Ct is the cell state vector.

After 3 gates, the current hidden state with the effective information will be outputted as

$$h_t = O_t \, tanh(C_t) \tag{19}$$

where ht is the hidden state.

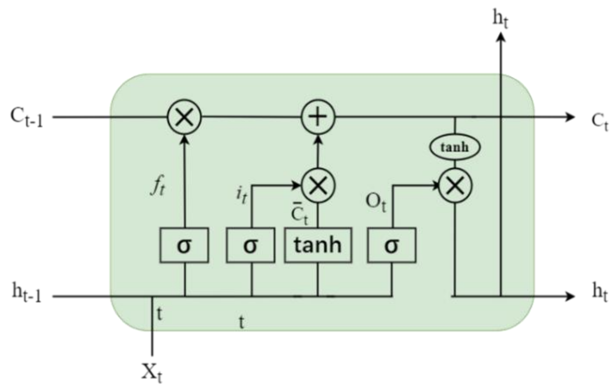

Figure 4 Structure of LSTM Memory Cell

3.9. Autoregressive Integrated Moving Average Model - Long Short-Term Memory model

Due to the coexistence of linear and non-linear relationships of the real data, neither ARIMA nor LSTM can satisfy the needs. Based on the features of these two models, a hybrid model is performed to make better predictions.

In the hybrid model, the ARIMA or SARIMA model deals with the linear information of the time series, and then the LSTM network processes the nonlinear part by analyzing the residuals of the ARIMA or SARIMA model.

3.10. Evaluation Indicator

Scientific and suitable evaluation indicators are applied to assess the prediction performance of candidate models. In this study, I used Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). The smaller the values of these indicators, the higher the goodness of fit the model.

1) Mean Squared Error (MSE):

MSE measures the number of errors in the forecasting models, calculating the average squared difference between the actual data and the estimators.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{x_i})^2 \tag{20}$$

where ei denotes the residual; xi denotes the actual observation; (x_i)^ denotes the predicted value.

2) Root Mean Squared Error (RMSE):

RMSE is the standard deviation of the residuals, which is the quadratic mean of the differences between the observations and the predictions. This indicator is defined as follows:

$$\text{RMSE}=\sqrt{MSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{x_i})^2} \tag{21}$$

where xi denotes the original observation; $\widehat{x_i}$ denotes the estimation of the prediction.

3) Mean Absolute Error (MAE):

MAE is calculated as the sum of absolute errors between the observations and predictions, which is defined as follows:

$$\text{MAE}=\frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y_i}| \tag{22}$$

4. DATA ANALYSIS

4.1. Data Description

Using data is about the number of confirmed cases of influenza in China every week from July 2009 to February 2024. The whole data set has 7 observations. The training set is from 5, July 2009, to 31, December 2023, where there are 757 observations. And the last 8 weeks data set as the validation set. The following Table 2 states its basic statistics.

Table 2 Basic Statistics of Raw Data

Mean	9833.948	Median	9833.948
Minimum	2269	Maximum	43201
Variance	48184659	Standard Deviation	6941.517

Skewness	2.310206	Kurtosis	8.746688

4.2. Stationarity Test

Figure 5 displays the time series plot of the number of confirmed influenza cases in China. Obviously, this time series has a positive trend and seasonality. Using the ADF test to verify, the p-value is 0.5528 and greater than α. Therefore, the null hypothesis cannot be rejected, indicating that this series is not stationary.

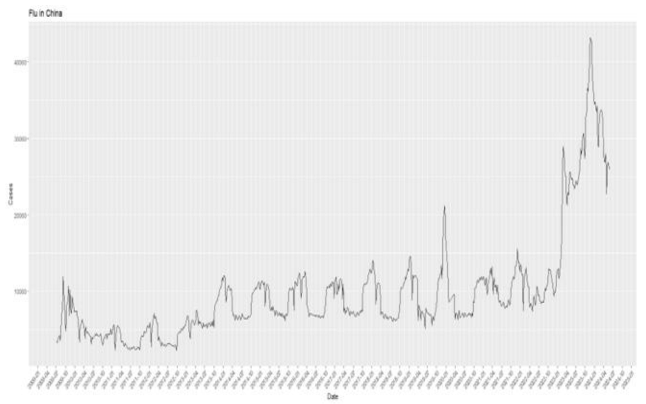

Figure 5 Time Series Plot of Raw Data

Since the stationarity is the precondition of the ARIMA model, differencing had to proceed. According to Figure 6, the time series plot of the first-order difference is much smoother. The new p-value of the ADF test changed to 0.01 so that the first difference series is stationary.

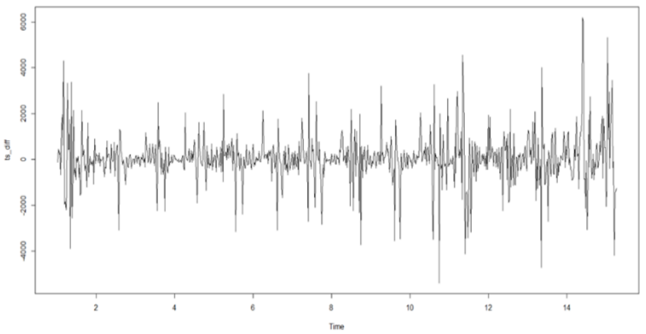

Figure 6 the Time Series Plot of 1st Difference data

4.3. ARIMA Model Establishment

Firstly, the order of difference (d) is definitely 1. Then, the ACF plot (Figure 7) has a tail-off, nevertheless, there is an apparent spike at lag1 in the PACF plot (Figure 8). Thus, the order of the AR component (p) and the order of the MA component (q) are possibly 0 and 1. On the other hand, in the long term, the ACF plot remains tail-off, meanwhile, the PACF has slow peaks at lag 52 and lag 104. Setting P = 0, Q = 2, and S = 52 for the seasonal parts is an option. In a word, the potential model is the ARIMA (1,1,0) (2,0,0)52 model.

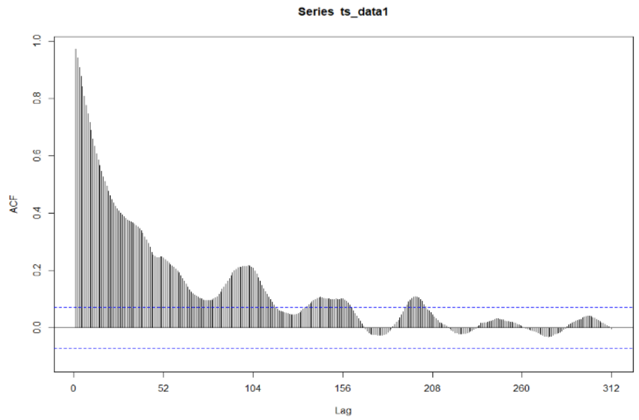

Figure 7 ACF plot

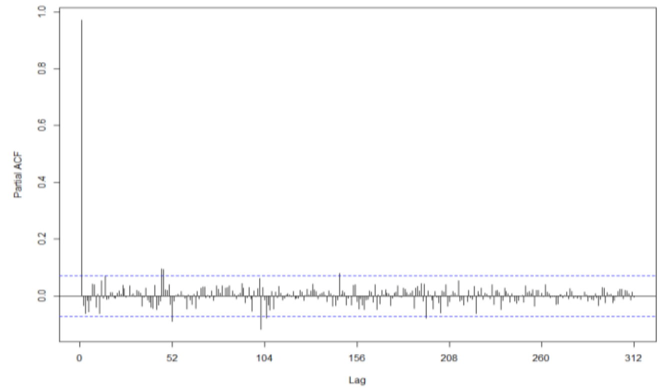

Figure 8 PACF Plot

The auto. arima function in R was applied to find the best ARIMA model by minimizing AIC and BIC and the SARIMA (0,1,0) (1,0,0)52 model was selected. Table 3 is the result of the estimation. The SAR1's p-value is approximately 0.0084 which states that we can reject the null hypothesis at the significant level of 5%. Hence, the SARIMA (0,1,0) (1,0,0)52 model is appropriate.

Table 3 Estimation of ARIMA Model

Parameter	SARIMA (0,1,0) (1,0,0)52	P value
SAR1	0.1137	0.008436995
MSE	1213855	
RMSE	1101.751	
MAE	660.4723	
Log-likelihood	-6251.14	
AIC	12506.31	
BIC	12515.51	

4.4. LSTM Model Establishment

In the hybrid model, the SARIMA model's residual became the input value to train the LSTM model. Figure 9 is the residual plot and we can see the range of their values vary widely so that the residuals were normalized by the Min-Max normalization to reduce the amount of computation and avoid overfitting in the training model.

Additionally, individual LSTM models were constructed using the original data to compare with the hybrid model. From Figure 5 and Table 2, we can see that the raw data varied greatly (Standard Deviation is 6941.517) and covered a wide range (from 2269 to 43201). Thus, normalization is even more necessary than the hybrid model.

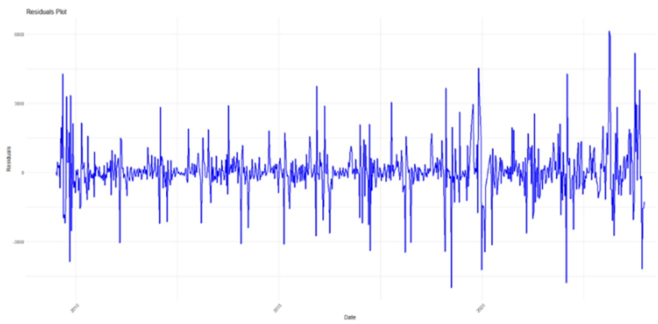

Figure 9 Residuals Plot

The architecture of the LSTM neural networks has 3 layers: one LSTM layer, one dropout layer, and one dense layer. There are many hyperparameters involved in the neural networks. On the one hand, some of them are fixed as shown in Table 3, based on relevant studies. Besides, we considered different numbers of epochs, such as 30, 40, 50, and 60, and batch sizes, such as 32 and 64 to get the optimal model.

Table 4 Hyperparameter of LSTM model

Hyperparameter	Value	Hyperparameter	Value
Optimizer	Adam	Activation function	Tach
Learning rate	0.001	Loss function	Mean Squared Error

The performance of implemented single LSTM models and the LSTM models of the combination model is shown in Table 4 and Table 5 severally where it can be seen that both LSTME models with 50 epochs and 32 batch sizes have the lowest RMSE and MAE. Therefore, their respective goodness of fit is the best. And they were selected to forecast.

Table 5 Errors of Individual LSTM Models

Epoch	Batch size	RMSE	MAE
30	32	0.0468147	0.0282226
	64	0.0568958	0.0343597
40	32	0.0324522	0.0197526
	64	0.0595546	0.0358315
50	32	0.0120231	0.0041258
	64	0.0542071	0.0326699
60	32	0.0221508	0.0132271
	64	0.0481036	0.0289136

Table 6 Errors of Hybrid LSTM Model

Epoch	Batch size	RMSE	MAE
30	32	0.0562290	0.0337770
	64	0.0610446	0.0366597
40	32	0.0368374	0.0221410
	64	0.0546519	0.0328715
50	32	0.0209959	0.0128318
	64	0.0531008	0.0318935
60	32	0.0221741	0.0137262
	64	0.0467891	0.0282501

4.5. Forecasting

The three fitted models above were applied to predict the value of the number of influenza cases next 8 weeks. Table 7 below shows the comparison of the predictions and the actual value and Figure 10 is their comparison plot to time. Apparently, the prediction of individual LSTM model is quite bad, there is a considerable gap between its predicted values and the validation set. The predictions of the other models were much better, their predicted values were close to the actual values, where the Hybrid model had the lowest residual.

Table 7 Actual Value vs. Predicted Value

Time step	Date	Actual Data	SARIMA (0,1,0)(1,0,0)52 Prediction	LSTM Prediction	Hybrid Prediction
1	2024/1/7	34773	34556.48	3458.369	34524.76
2	2024/1/14	34339	34504.49	3340.993	34378.00
3	2024/1/21	33501	34112.23	3774.985	34421.77
4	2024/1/28	34218	34429.07	3981.230	34556.93
5	2024/2/4	29206	34491.01	4287.124	34696.98
6	2024/2/11	28927	35488.62	3636.151	34951.16
7	2024/2/18	33038	34547.91	5064.019	35654.94
8	2024/2/25	33280	34848.27	28446.38	36340.30

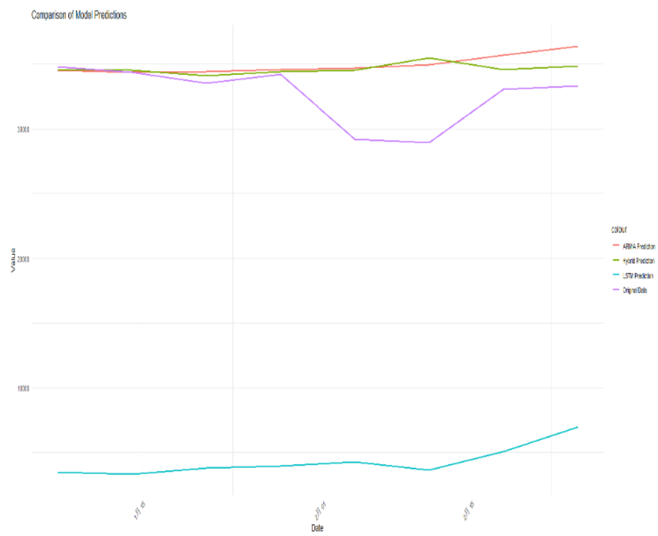

Figure 10 Compared Time Series Plot of Actual Value vs. Predicted Value

Table 8 shows the forecasting performance of every model. Certainly, all of the values of evaluating indicators of the single LSTM model are very large. The hybrid displayed the best performance with an approximate reduction of 4.6% of RMSE, 8.9% of MSE, and 13.9% of MAE. And we can also see from Figure 10 that the ARIMA model forecasting the trend was better than the individual LSTM in the early stage, but its residuals increased more and more over time. Conversely, the errors of the individual LSTM model had a decreasing trend in the later period. These reflected the models' features: the ARIMA model is good at predicting short-term trends and the LSTM model can detect long-term trends. The joint model integrated their advantages to improve the accuracy of predictions.

In addition, the poor performance of the single LSTM may be due to insufficient sample size. Constantin and Gyorgy (2024) proposed that over-fitting situations more likely to appear with small sample sizes. But this phenomenon didn't occur in the new ARIMA-LSTM model shows that the incorporation of the ARIMA and the LSTM methods indeed can overcome their weaknesses

Table 8 Performance of Models Forecasting

Date	SARIMA (0,1,0) (1,0,0)52	LSTM	Hybrid
RMSE	3234.213	809196603	3086.62
MSE	10460132	980170449	9527222
MAE	2342.416	28343.420	2016.138

5. Conclusion AND FUTURE WORK

This paper explores how combining the ARIMA and LSTM neural networks can improve time series analysis predictions. The ARIMA model and the LSTM are commonly used in sequential and time series data because they can capture linear and nonlinear information, respectively. In the hybrid model, the ARIMA model's residuals are passed into the LSTM network, effectively combining their strengths. The establishment of the integrated model is relatively easy to achieve. Training LSTM networks requires a lot of computing and has fairly high requirements for hardware facilities and data. In practical problems such as influenza, this shortcoming is magnified, which is proved by the poor predictive effect of the individual LSTM model in this study. Reducing the input value to the residual effectually improves this problem. Besides, the decrease in RMSE, MSE, and MAE in the hybrid model indicates that further residual analysis can compensate for the insensitivity of the ARIMA model to nonlinear relationships. Therefore, utilizing a hybrid model can not only enhance prediction accuracy but also make implementation easier.

On the other hand, this paper affords a useful tool to forecast the number of influenza cases in China. This model could provide support to decision-makers, such as the government, by detecting the spread of flu in advance to prevent the outbreak of an epidemic.

The outcome of this research also contributes to the literature by extending this type of technology combination into the influenza field. And the improvement in predicted accuracy demonstrated the previous researchers' proposition.

However, the amount of data used in this study is limited. The hybrid model was probably supposed to overfit, although it is avoided in this model. There is still the risk of overfitting when we will apply this method to other cases.

In the future, we would like to extend this algorithm to other fields, further verify its operability, and improve the construction of hybrid models to further improve the accuracy of predictions, such as introducing other variables. And we will try to fix the structure of the LSTM model including the hybrid model's LSTM part to fix the problem of overfitting.

References

[1] Box, G.E. and Tiao, G.C. Intervention Analysis with Applications to Economic and Environmental Problems. Journal of the American Statistical Association, 1975, 70: 70-79.

[2] Alan Lapedes and Robert Farber. How neural nets work. In Proceedings of the 1987 International Conference on Neural Information Processing Systems (NIPS'87), 1987, pp. 442–456. Cambridge, MA, USA : MIT Press

[3] Kanad Chakraborty, Kishan Mehrotra, Chilukuri K. Mohan, Sanjay Ranka. Forecasting the behavior of multivariate time series using neural networks. Neural Networks, 1992, 5:961-970.

[4] Setyo Tri Wahyudi. The ARIMA Model for the Indonesia Stock Price. Wahyudi2017TheAM. 2017.

[5] Emmanuel Dave, Albert Leonardo, Marethia Jeanice, Novita Hanafiah. Forecasting Indonesia Exports using a Hybrid Model ARIMA-LSTM [J]. Procedia Computer Science, 2021, 179: 480-484.

[6] Wang YW, Shen ZZ, Jiang Y. Comparison of ARIMA and GM (1,1) models for prediction of hepatitis B in China [J]. PLoS One, 2018, 13(9): e0201987.

[7] Ospina, R.; Gondim, J.A.M.; Leiva, V.; Castro, C. An Overview of Forecast Analysis with ARIMA Models during the COVID-19 Pandemic: Methodology and Case Study in Brazil [J]. Mathematics, 2023, 11: 3069.

[8] Sima Siami-Namini & Akbar Siami Namin. Forecasting Economics and Financial Time Series: ARIMA vs. LSTM [J]. ArXiv.org, 2018, abs/1803.06386.

[9] Ashutosh Kumar Dubey, Abhishek Kumar, Vicente García-Díaz, Arpit Kumar Sharma, Kishan Kanhaiya. Study and analysis of SARIMA and LSTM in forecasting time series data [J]. Sustainable Energy Technologies and Assessments, 2021, 47: 10147.

[10] Wu DCW, Ji L, He K, Tso KFG. Forecasting Tourist Daily Arrivals With A Hybrid SARIMA–LSTM Approach [J]. Journal of Hospitality & Tourism Research, 2021, 45: 52-67.

[11] Divino JA, McAleer M. Modelling and forecasting daily international mass tourism to Peru [J]. Tourism Management, 2010, 31: 846-854.

[12] Sherstinsky A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network [J]. Physica D: Nonlinear Phenomena, 2020, 404: 132306.

[13] Stanford JL, Vardeman SB. Random Process Models [J]. Methods in Experimental Physics, C. Chatfield, 1994, 28: 63-92.

[14] Kotu V, Deshpande B. Time Series Forecasting [J]. Data Science, 2nd ed., Vijay Kotu, Bala Deshpande, 2019, pp. 395-445.

[15] Tealab A, Hefny H, Badr A. Forecasting of nonlinear time series using ANN [J]. Future Computing and Informatics Journal, 2017, 2: 39-47.

[16] Sahraei A, Chamorro A, Kraft P, Breuer L. Application of Machine Learning Models to Predict Maximum Event Water Fractions in Streamflow [J]. Frontiers in Water, 2021, 3.

[17] Ludwig SA. Comparison of Time Series Approaches Applied to Greenhouse Gas Analysis: ANFIS, RNN, and LSTM [J]. IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2019, pp. 1-6, New Orleans, LA, USA.

[18] Selvin S, Vinayakumar R, Gopalakrishnan EA, Menon VK, Soman KP. Stock Price Prediction Using LSTM, RNN and CNN-Sliding Window Model [J]. International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017, pp. 1643-1647, Udupi, India.

[19] Staudemeyer RC, Rothstein Morris E. Understanding LSTM - A Tutorial Into Long Short-Term Memory Recurrent Neural Networks [J]. ArXiv, 2019, abs/1909.09586.

[20] Kim TK, Park JH. More About the Basic Assumptions of T-Test: Normality and Sample Size [J]. Korean Journal of Anesthesiology, 2019, 72: 331-335.

[21] Ford C. Understanding Robust Standard Errors [J]. UVA Library StatLab, 2020.

[22] Shapiro SS, Wilk MB. An Analysis of Variance Test for Normality (Complete Samples) [J]. Biometrika, 1965, 52: 591–611.

[23] Mushtaq R. Augmented Dickey Fuller Test [J]. Econometrics: Mathematical Methods & Programming eJournal, 2011.

[24] Wikipedia contributors. Augmented Dickey-Fuller Test [J]. Wikipedia, The Free Encyclopedia, 2024.

[25] Razali NM, Yap BW. Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests [J].

[26] Cryer JD, Chan KS. Time Series Analysis with Applications in R (2nd ed.) [J]. New York: Springer, 2008.

[27] Yang ZR, Yang Z. Artificial Neural Networks [J]. Comprehensive Biomedical Physics, Anders Brahme, 2014, pp. 1-17, Elsevier.

[28] Hamami F, Dahlan IA. Univariate Time Series Data Forecasting of Air Pollution Using LSTM Neural Network [J]. International Conference on Advancement in Data Science, E-learning and Information Systems (ICADEIS), 2020, pp. 1-5, Lombok, Indonesia.

[29] Pooniwala N, Sutar R. Forecasting Short-Term Electric Load with a Hybrid of ARIMA Model and LSTM Network [J]. International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-6, Coimbatore, India.

[30] Venna SR, Tavanaei A, Gottumukkala RN, Raghavan VVR, Maida AS, Nichols S. A Novel Data-Driven Model for Real-Time Influenza Forecasting [J]. IEEE Access, 2019, 7: 7691-7700.

[31] Wen X, Li W. Time Series Prediction Based on LSTM-Attention-LSTM Model [J]. IEEE Access, 2023, 11: 48322-48331.

[32] Kandula S, Shaman J. Near-Term Forecasts of Influenza-Like Illness: An Evaluation of Autoregressive Time Series Approaches [J]. Epidemics, 2019, 27: 41-51.

[33] Tsan Y-T, Chen D-Y, Liu P-Y, Kristiani E, Nguyen KLP, Yang C-T. The Prediction of Influenza-Like Illness and Respiratory Disease Using LSTM and ARIMA [J]. International Journal of Environmental Research and Public Health, 2022, 19: 1858.

[34] Al-Qaness MAA, Ewees AA, Fan H, Elaziz MA. Optimized Forecasting Method for Weekly Influenza Confirmed Cases [J]. International Journal of Environmental Research and Public Health, 2020, 17: 3510.

[35] Azad AS, Sokkalingam R, Daud H, Adhikary SK, Khurshid H, Mazlan SN A, Rabbani MBA. Water Level Prediction through Hybrid SARIMA and ANN Models Based on Time Series Analysis: Red Hills Reservoir Case Study [J]. Sustainability, 2022, 14: 1843.