Research and Application of Time Series Forecasting Methods in Disease Surveillance

Xue Bai[1] and Yang Yu[1]

College of Artificial Intelligence, Shenyang Normal University
110034 Shenyang, China
Corresponding author: Yang Yu *Received Nov. 14, 2025; Revised and Accepted Dec. 24, 2025*

Abstract. Time series forecasting plays a pivotal role in public health, particularly within disease surveillance. Theoretically, it facilitates the in-depth analysis of disease patterns and progression; practically, it supports health authorities in optimizing resource allocation, devising prevention strategies, and mitigating transmission risks. Recently, rapid advancements in information technology have accelerated the development of this field, leading to the continuous emergence of novel methods. This paper provides a comprehensive overview of time series forecasting methods, categorizing them into traditional statistical models and deep learning approaches, while strictly evaluating their respective strengths and limitations. Furthermore, it explores innovative applications involving statistical models, machine learning, and deep learning, highlighting the potential of these emerging technologies. Bridging theory and practice, this systematic review aims to establish an analytical framework, delineate the field's developmental trajectory, and offer valuable insights for future research in disease prediction.

Keywords: Time Series Forecasting, Disease Surveillance, Deep Learning, Transformer.

1. Introduction

In an era where the global public health security architecture faces multifaceted threats, the role of disease surveillance as a sentinel for disease prevention and control is being prioritized with unprecedented urgency. Recently, the landscape of disease epidemiology has become increasingly intricate. The spectrum ranges from the cyclical resurgence of traditional pathogens, such as seasonal influenza, to the abrupt emergence of novel infectious diseases like COVID-19, presenting the global public health system with unparalleled challenges [1]. Data from the World Health Organization (WHO) indicate that while mortality from infectious diseases remains critically high, the escalating prevalence of chronic diseases is also placing a substantial strain on societal medical resources. Consequently, capturing disease dynamics with precision and forecasting developmental trends have emerged as paramount requirements for optimizing public health emergency responses and devising evidence-based control strategies [2].

Time series forecasting, a pivotal branch of data analysis, holds significant promise for disease surveillance due to its capacity to extract profound insights from historical dynamic patterns. The field has witnessed a continuous evolution of methodologies, spanning from traditional Autoregressive Integrated Moving Average (ARIMA) models to machine learning-integrated approaches like Random Forest and Support Vector Machines (SVM), and ultimately to advanced deep learning architectures such as Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), and Transformers. Traditional ARIMA models excel at capturing linear trends, seasonality, and periodicity in stationary or difference-stationary time series, making them widely applicable in forecasting epidemics like influenza and hand, foot, and mouth disease [3,4]. In contrast, machine learning algorithms have emerged as powerful tools for addressing non-linear and high-dimensional disease data; specifically, Random Forest is adept at feature selection and trend prediction [5], while SVMs facilitate precise classification and regression regarding disease severity and occurrence [6]. Most notably, deep learning models have driven a paradigm shift in disease prediction through their superior feature learning capabilities. LSTM and GRU networks demonstrate remarkable efficacy in handling long-term dependencies within infectious disease trends [7], whereas Transformers, utilizing self-attention mechanisms, deliver exceptional performance in modeling dependencies across long-sequence data.

While time series forecasting has yielded promising results in disease surveillance, its real-world implementation faces substantial obstacles. The inherent complexity of disease data, manifesting as incompleteness, noise, non-stationarity, and non-linearity, places rigorous demands on model adaptability and precision. Furthermore, limitations regarding model interpretability, computational efficiency, and the integration of multi-source heterogeneous data impede the broader deployment and sophistication of these forecasting techniques. Consequently, a systematic review of the field's evolution, a critical examination of the merits and demerits of existing models, and an exploration of strategies to address data complexity are of paramount theoretical and practical importance.

Such efforts are essential for optimizing disease surveillance and ensuring global public health security. By conducting a comprehensive survey of these pivotal issues, this paper aims to offer significant insights to guide future research and practice in public health

2. Overview of Time Series Forecasting

2.1. Time Series Data Characteristics

Time series data is defined as a collection of data points organized chronologically, where each point is indexed to a specific time. These data may be sampled at fixed frequencies or irregular intervals, yet they share a common attribute of strict sequentiality. Generally, each data instance comprises a timestamp accompanied by one or more univariate or multivariate observations. Such data play a pivotal role in domains ranging from financial markets and meteorology to transportation logistics and industrial manufacturing. Formally, a time series is denoted as $X = x(t_1), x(t_2), \cdots, x(t_n)$, where $t_1 < t_2 < \cdots < t_n$ serves as the time index and $x(t_i)$ represents the observation at moment t_i. A defining characteristic of time series is its temporal ordering, which encodes essential information; any permutation of this order would irrevocably alter the semantic interpretation of the data. Time series forecasting, therefore, entails the use of mathematical models to predict future values by leveraging intrinsic properties and latent patterns mined from historical sequences. An analysis of the specific characteristics of time series data follows.

1. **Temporal Dependency:** Time series observations adhere to a strict chronological sequence, where each data point is contingent upon its predecessors. This dependency imparts a ”memory” characteristic to the data, implying that current states are functions of historical values. Consequently, models must explicitly incorporate these temporal linkages to effectively capture dynamic evolution.
2. **Trend:** Data may exhibit persistent long-term upward or downward trajectories, manifesting as directional shifts over time. This attribute reflects the underlying developmental tendency of the series. Proper identification and processing of trends are critical to prevent models from overfitting to specific segments or neglecting global patterns.
3. **Seasonality:** Time series frequently demonstrate cyclical fluctuationsłrecurring patterns observed across annual, quarterly, monthly, weekly, or daily intervals. These periodicities represent repetitive variations intrinsic to specific temporal windows.
4. **Stochasticity:** The data may comprise random fluctuations or noise attributable to neither trends, seasonality, nor other deterministic factors. Such stochasticity introduces uncertainty, complicating the forecasting task. Mitigation strategies, including smoothing algorithms or the integration of stochastic components, are necessary to manage this irregularity.
5. **Non-stationarity:** Statistical properties such as mean and variance may evolve over time, defying the assumption of stationarity. As traditional statistical methodologies are often predicated on stationary data, non-stationarity necessitates pre-processing stepsłsuch as differencing or logarithmic transformationsłto induce stationarity.
6. **Correlation:** Interdependencies often exist among variables within time series data, where variations in one covariate precipitate changes in others. This correlation enriches the information available for modeling. Leveraging multivariate frameworks that incorporate these correlated variables can significantly enhance predictive accuracy.
7. **High-dimensionality and Complexity:** Real-world time series often encompass multiple variables spanning diverse temporal scales, resulting in high dimensionality and structural complexity. These factors exacerbate analytical and modeling challenges. Advanced techniques, such as feature selection and dimensionality reduction, are pivotal in reducing data complexity and improving computational feasibility.

2.2. Evaluation Metrics

Evaluation metrics are essential instruments for quantifying model efficacy, offering an objective foundation for analysis. Distinct metrics elucidate model characteristics from diverse angles, making them applicable to varying research contexts. This section delineates the evaluation metrics frequently utilized in time series forecasting, examining their theoretical principles, distinct features, and value in practical applications.

1. Mean Squared Error (MSE). As a standard metric for assessing the deviation between predictions and ground truth, MSE computes the mean of the squared differences. Its sensitivity to large errors allows it to effectively penalize severe outliers. However, the squaring operation renders the metric’s units inconsistent with the

original data, thereby diminishing interpretability. MSE is optimal for scenarios where minimizing the overall error magnitude is critical and outlier sensitivity is desired. It is formally defined as Equation (1):

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

where y_i is the actual value of the $i-th$ sample, $\hat{y}_i$ is the predicted value of the $i-th$ sample, and n represents the total sample size.

2. Mean Absolute Error (MAE). MAE quantifies the average magnitude of absolute errors. Unlike MSE, its dimensionality aligns with the original data. The use of absolute values neutralizes the effect of error direction, ensuring non-negativity. MAE does not amplify errors, rendering it highly robust to outliers. Conversely, its inability to indicate error direction may mask local biases. It is ideal for forecasting tasks demanding stable and interpretable error evaluation. It is defined as Equation (2):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|. \tag{2}$$

3. Mean Absolute Percentage Error (MAPE). MAPE derives a percentage-based index of overall predictive performance by averaging the ratio of absolute errors to actual values. Being a dimensionless metric, it facilitates comparisons across datasets with disparate scales or units and offers an intuitive representation of relative error. However, it is susceptible to instability when actual values are near zero, potentially leading to undefined results. MAPE is particularly advantageous for tasks necessitating relative error assessment. It is defined as Equation (3):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{y_i - \hat{y}_i}{y_i}| \times 100\%. \tag{3}$$

3. Classification of Time Series Forecasting Models

3.1. Traditional statistical models

Traditional statistical models, predicated on established statistical theories, facilitate forecasting by extracting statistical regularities and patterns from time series data. Distinguished by their solid theoretical basis and computational efficiency, these models are ubiquitous in applications characterized by limited data volume or well-defined features. Generally, they postulate that data follow linear dynamics and employ parameter estimation alongside model order determination to fit historical sequences, thereby generating predictions for future time points.

Autoregressive Integrated Moving Average Model (ARIMA) The Autoregressive Integrated Moving Average (ARIMA) model, also referred to as the Box-Jenkins model, is a seminal time series forecasting methodology introduced by Box and Jenkins [8]. Fundamentally, the model operates by transforming non-stationary time series into stationary sequences, subsequently modeling them via a synthesis of Autoregressive (AR) and Moving Average (MA) processes. The model is formally denoted as ARIMA(p,d,q), and its architectural workflow is depicted in Figure 1. In this notation, p signifies the autoregressive order, capturing the linear correlation between the current observation and its p lagged values; d denotes the degree of differencing required to induce stationarity; and q represents the moving average order, modeling the relationship between the current value and the forecast errors of the preceding q periods. The AR component leverages the linear dependence of the current value on its historical observations for prediction. Conversely, the MA component predicts future values based on historical error terms, positing that the current value is a linear combination of past errors. This component is particularly effective at capturing short-term fluctuations and describing the direct influence of white noise (random shocks). Essentially, ARIMA extends the Autoregressive Moving Average (ARMA) model by integrating a differencing mechanism to handle non-stationary data.

Faced with time series data characterized by pronounced seasonality, the Seasonal Autoregressive Integrated Moving Average (SARIMA) model extends the traditional ARIMA framework by explicitly integrating seasonal autoregressive, seasonal differencing, and seasonal moving average components. This architecture is specifically designed for data manifesting clear seasonal periodicity. The model is mathematically represented as $SARIMA(p, d, q)(P, D, Q)s$. In this notation, p, d, and q correspond to the orders of the non-seasonal autoregressive, differencing, and moving average terms, respectively. Analogously, P, D, and Q signify the orders of the seasonal autoregressive, seasonal differencing, and seasonal moving average terms, while s represents the periodicity of the seasonal cycle.Several traditional statistical models are compared in Table 1.

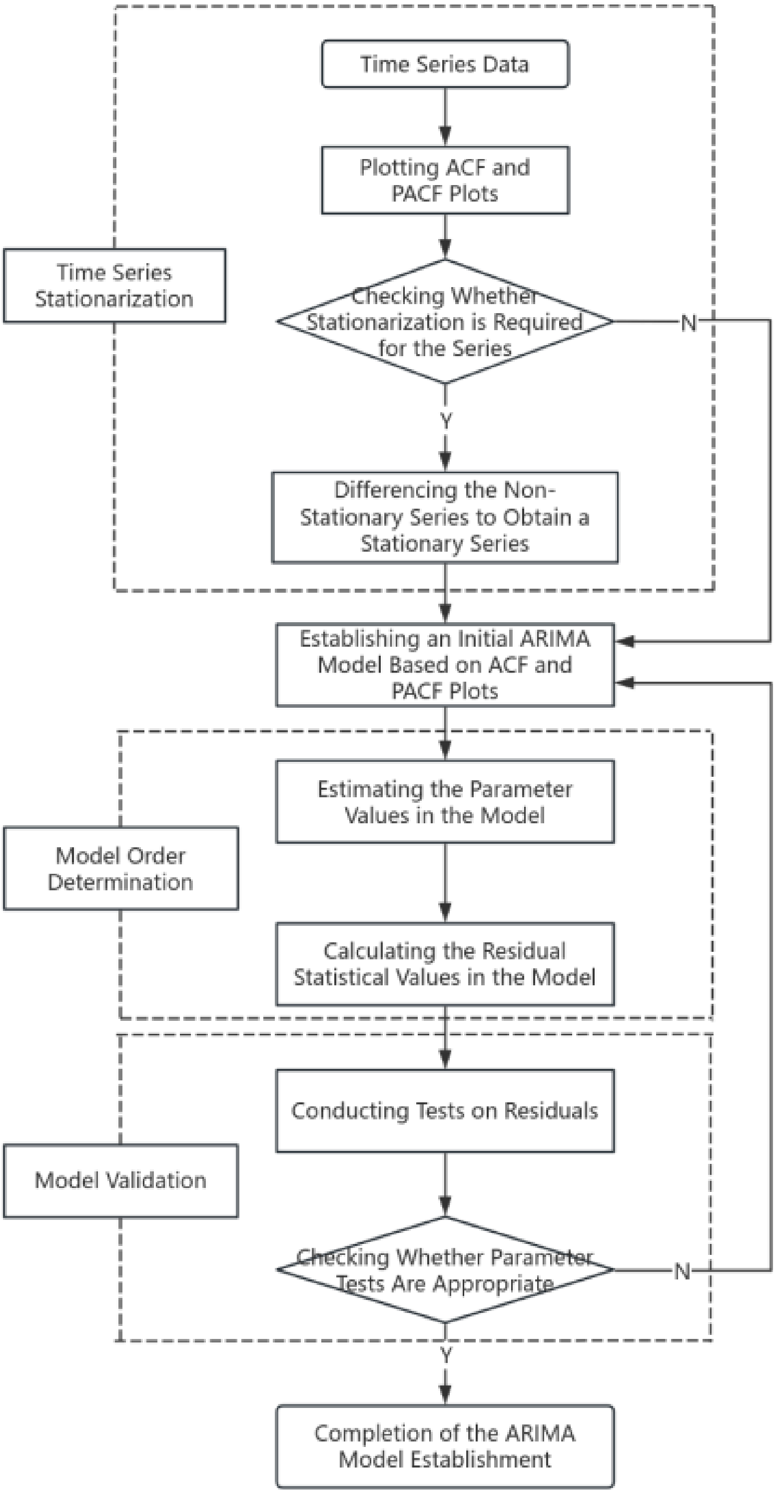

Fig. 1. Construction Process of the ARIMA Model

Table 1. Comparison of Traditional Statistical Models

Model	Core Idea	Predictive Capability
AR	Current values are determined by a linear combination of the previous p-period values	Long-term data.
MA	Current values are determined by a linear combination of random errors in the previous q periods	Can only predict within q steps; long-term prediction degenerates to the mean.
ARMA	Combines AR and MA, considering the effects of both historical values and historical errors simultaneously	Superior to single AR/MA models, suitable for short-to-medium-term prediction.
ARIMA	Incorporates differencing into ARMA to handle trendiness.	Can capture trend changes, suitable for medium-to-long-term prediction.
SARIMA	Adds seasonal parameters to ARIMA to handle periodic fluctuations	Can capture both trends and seasonal cycles simultaneously.

The ARIMA model, a staple in time series analysis, plays a pivotal role in disease surveillance research, offering scientific evidence to guide early warning systems and prevention strategies. A variety of studies have leveraged this model to forecast diverse pathologies. For instance, Cai et al. [9] modeled Hepatitis B trends using data from Zhabei District, Shanghai. By establishing an ARIMA(1,1,1) model, they observed residual autocorrelation coefficients within 0.5 and a strong concordance between predicted and observed values, demonstrating the model's efficacy. Similarly, Zheng and Zhai [10] analyzed Class B infectious disease rates in Shenzhen (2006C2019) using an ARIMA(2,1,1) framework. Their model demonstrated superior performance, capturing short-term trends with actual values consistently falling within the 95% confidence interval. In another study, Wu et al. [11] applied an ARIMA(0,1,2)(0,1,1)12 model to Hepatitis B data from Jiangsu Province (2000C2009). The model yielded a white noise residual sequence and a mean relative error of 8.26%, providing critical insights for adjusting provincial control policies. Furthermore, Zheng et al. [12] focused on foodborne diseases in Baotou City (2017C2020), constructing an ARIMA(0,1,1)(0,1,1)12 model. With a BIC of 9.218 and a non-significant Ljung-Box Q-test indicating white noise residuals, the model successfully captured the disease's incidence trends. Collectively, these studies encompass a wide range of geographical regions, time periods, and disease categoriesłincluding Hepatitis B, Class B infectious diseases, and foodborne illnessesłhighlighting the adaptability of ARIMA modeling through specific parameterization for distinct research targets.

Exponential Smoothing Model The Exponential Smoothing model is a prominent statistical technique in time series forecasting. Fundamentally, it estimates future values by computing a weighted average of historical observations, assigning exponentially diminishing weights to older datałthereby placing greater emphasis on recent observations to mitigate random fluctuations. This methodology is particularly effective for data demonstrating trends, seasonality, and time dependence. Depending on the specific trend and seasonal characteristics inherent in the data, the approach is categorized into Simple, Double, and Triple Exponential Smoothing. Among these, Simple Exponential Smoothing (SES) is specifically designed for stationary time series devoid of trend or seasonal components. The mathematical formulation is defined as Equation (4).

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t. \tag{4}$$

In this equation, $\hat{y}_{t+1}$ denotes the forecast for time $t+1$, y_t is the actual observation at time t, $\hat{y}_t$ represents the forecast for time t, and α (where $0 < \alpha < 1$) serves as the smoothing parameter.

Double Exponential Smoothing (DES), commonly referred to as Holt's Linear Trend method, is specifically designed for time series characterized by a linear trend in the absence of seasonality. By integrating a trend component into the SES architecture, it effectively captures the data's long-term evolutionary trajectory. Triple Exponential Smoothing (TES), known as the Holt-Winters method, extends this approach to time series exhibiting both trend and seasonal dynamics. It augments the DES model with a seasonal component to account for periodic fluctuations. Through the calibration of smoothing parameters, these models optimize the balance between recent observations and historical data, thereby enhancing predictive accuracy and stability. Moreover, the exponential smoothing framework can be generalized to accommodate intricate patterns, including non-linear trends and complex seasonalities.

Research into the application of Exponential Smoothing (ES) for disease surveillance has yielded significant findings. Wang Xin et al. [13] highlighted that ES models effectively capture short-term trends in Influenza-like

Illness (ILI), proving particularly robust for data exhibiting minimal temporal volatility. In a study targeting Hepatitis A trends in Shanghai, Zhu Yiyi et al. [14] leveraged both Holts linear trend and Holt-Winters multiplicative models, demonstrating that these methods maintain high predictive accuracy across varying data types. Comparatively, Liu Tian et al. [15] evaluated SARIMA against ES models for forecasting Mumps incidence, revealing that ES possesses superior adaptability in processing time series data. Similarly, Tang Guangxin et al. [16] validated the efficacy of ES in infectious disease forecasting using measles data from 2005 to 2015. Furthermore, in an analysis of Schistosomiasis control in Hunan Province, Zhou Xiaonong et al. [17] utilized ES and ARIMA models to project human and livestock infection rates, finding that ES offers distinct advantages in low-prevalence scenarios. These studies collectively underscore the extensive potential of exponential smoothing in disease surveillance, particularly its adaptability and precision in handling time series data and generating short-term forecasts.

3.2. Deep Learning Model

Driven by rapid advancements in artificial intelligence, deep learning models have exhibited formidable expressive power in time series forecasting. Grounded in neural network architectures, these models possess the intrinsic capability to automatically extract intricate non-linear features and long-term dependencies from large-scale data, making them particularly adept at handling high-dimensional and unstructured time series.

Long Short-Term Memory Network (LSTM) Long Short-Term Memory (LSTM) represents a specialized class of Recurrent Neural Networks (RNNs) engineered to process sequential data, including time series and natural language. Central to the LSTM architecture is its gating mechanism, consisting of input, forget, and output gates. These gates modulate information flow, granting the model the capacity to selectively preserve or discard information, which is critical for capturing long-term dependencies. In the context of time series forecasting, LSTM mitigates the vanishing and exploding gradient issues that plague traditional RNNs by leveraging Memory Cells and gated structures. This design ensures robust performance when processing long sequences, as depicted in Figure 2. Moreover, LSTM excels at managing noisy and highly volatile data while modeling features across diverse temporal scales, thereby significantly improving both predictive accuracy and model robustness.

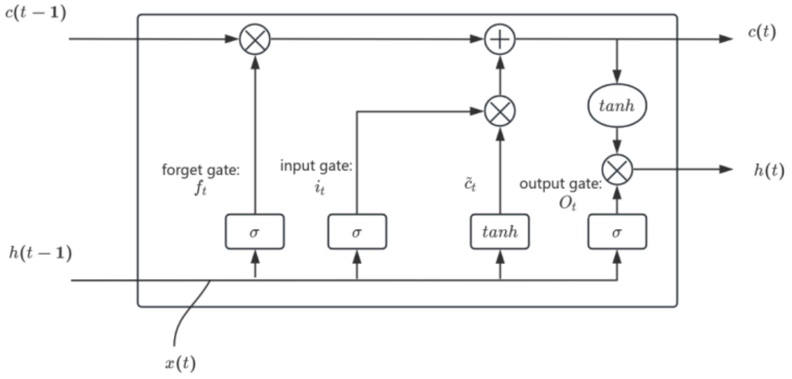

Fig. 2. LSTM Architectural Diagram

Due to their capacity to capture long-term dependencies in time series, LSTM networks have been extensively adopted for modeling infectious disease transmission. These models facilitate the simulation of spatiotemporal propagation processes by incorporating variables such as infection sources, meteorological conditions, diffusion patterns, and intervention strategies to assess epidemic risks. In the context of the COVID-19 pandemic, Bharadi et al. [18] underscored the critical role of accurate forecasting in public health governance. They elucidated the mechanisms of Stateful versus Stateless LSTM modes, demonstrating through empirical experiments that the Stateful mode yields superior accuracy and generalization, particularly for extended time series. Conversely, while the Stateless mode offers computational efficiency, it lacks precision in long-term trend forecasting. The study concluded by advocating for the Stateful approach and proposing future optimization strategies. Similarly, Mukhtar et al. [19] validated the generalization capability of LSTM on complex, non-stationary epidemic data. Leveraging multivariate inputsłincluding case growth rates and population mobility indicesłalongside an adaptive time-step design, their model achieved high-precision forecasts for short-term (7-day) COVID-19 trends, demonstrating robustness even in the absence of complete historical records.

In the realm of disease monitoring and prognosis, LSTM also demonstrates remarkable efficacy. It enables the analysis of physiological parametersłsuch as heart rate, blood pressure, and glucose levelsłto facilitate the timely detection of latent health issues. Li et al. [20] utilized LSTM to handle irregularly sampled time series within Electronic Health Records (EHR), targeting the prediction of 30-day readmission risks for heart failure patients. By learning long-term dependencies of key physiological indicators via gating mechanisms, the model achieved an AUC of 0.89, outperforming logistic regression models by 12%. This study corroborates LSTM's ability to mine the dynamic evolutionary patterns of disease pathology for early intervention. Furthermore, Lee and Hauskrecht [21] introduced a dual-information fusion LSTM model that synthesizes recent clinical events with historical hidden states. By integrating immediate observations with long-term temporal features, the model optimized predictions for clinical outcomes, such as disease onset and treatment response. Experiments on the MIMIC-III intensive care dataset revealed that this approach significantly surpasses models relying on single information sources, offering a novel framework for personalized, dynamic disease risk monitoring.

Transformer Model and Its Derivatives The Transformer model operates on the core principles of self-attention and multi-head attention mechanisms, designed to capture global dependencies across input sequences and enable parallelized processing. Diverging from traditional Recurrent Neural Networks (RNNs) or Convolutional Neural Networks (CNNs), the Transformer relies solely on attention mechanisms, eschewing recurrent architectures and convolutional operations. This design significantly enhances both computational efficiency and overall model performance. For time series forecasting, the model utilizes an encoder-decoder architecture to explicitly model long-term dependencies through self-attention, thereby improving predictive precision. Its architectural schematic is presented in Figure 3. Moreover, the Transformer is adept at handling multivariate time series data and employs positional embeddings to encode the inherent sequential order of the data.

Leveraging their powerful global modeling capabilities, Transformer models have become a focal point of research in time series forecasting. Notable advancements include the Informer model, which mitigates the computational complexity and memory constraints of long-sequence prediction through the introduction of ProbSparse self-attention and a generative decoder. Similarly, the Autoformer model utilizes a specific Decomposition Layer to disentangle series data, thereby enhancing the modeling of seasonal and trend components. Additionally, the MetaTST model incorporates textual metadata into the Transformer framework, significantly augmenting predictive interpretability and precision. These advanced Transformer variants have exhibited superior performance in diverse applications, ranging from power consumption and traffic flow analysis to weather forecasting. Table 2 outlines the potential optimization directions and distinct advantages of these Transformer derivatives specifically for disease surveillance.

Wu et al. [26] pioneered the application of the standard Transformer architecture to epidemic forecasting, specifically targeting Influenza-like Illness (ILI) incidence data. By leveraging self-attention mechanisms, the model effectively learns non-linear dependencies embedded in time series. Capable of processing both univariate (case counts) and multivariate (integrating temperature and humidity) inputs, it outperformed traditional ARIMA and LSTM models in multi-step forecasting of U.S. CDC ILI data, achieving an RMSE of 0.83. These results validated the Transformer's universal capacity to capture long-range temporal dynamics, thereby establishing a robust baseline for future research in medical time series analysis.

Furthermore, Meryem Altin Karagozdeb et al. [27] conducted a systematic evaluation of Transformer variants, such as Informer, Autoformer, and PatchTST, for diabetes blood glucose prediction. The study integrated Continuous Glucose Monitoring (CGM) data with dietary records as inputs. Empirical results revealed that the seasonal decomposition module inherent to the Autoformer significantly enhanced the prediction of periodic glucose fluctuations, yielding a 23% reduction in MAE. Conversely, the channel-independence strategy employed by PatchTST improved adaptability to inter-patient variability, offering a novel paradigm for personalized modeling in chronic disease management.

4. Basic Ideas and Steps for the Establishment of Time Series Forecasting Models

In the realm of disease surveillance, time series forecasting leverages historical temporal patterns to project future disease trends and transmission scales. The fundamental strategy involves integrating epidemiological characteristics with advanced time series analysis to construct high-precision predictive models. The methodological framework comprises the following steps:

(1) Data Acquisition and Preprocessing: This step involves aggregating multi-dimensional surveillance datałcovering incidence, mortality, and outpatient volumełwhile synthesizing relevant covariates such as demographics,

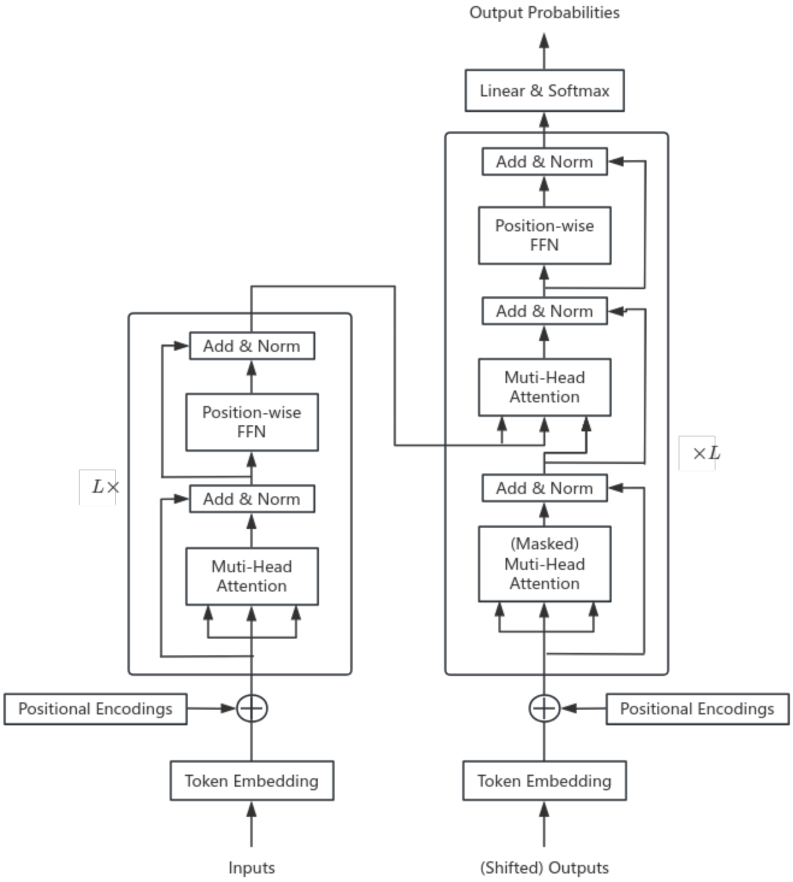

Fig. 3. Architectural Diagram of the Transformer Model

Table 2. Comparison of the Transformer and Its Derivatives

Model Type	Representative Model	Innovative Mechanism	Advantages in Disease
Long Sequence Optimization	Informer [22]	ProbSparse Attention + Generative Decoder	Processes long-cycle data of epidemic spread
Temporal Decomposition	Autoformer [23]	Sequence Decomposition Embedded in Encoder-Decoder	Captures seasonal fluctuations of chronic diseases
Multimodal Fusion	MetaTST [24]	Metadata Embedding + Multi-source Feature Alignment	Integrates multi-factors to predict infectious disease risks
Lightweight Design	LDTransformer [25]	Compressed Transformer Layers + Knowledge Distillation	Applicable to resource-constrained scenarios

meteorological factors, and holiday effects. To address challenges like underreporting and misreporting, data integrity must be rigorously verified. Missing values are imputed using statistical techniques (e.g., mean imputation based on spatiotemporal similarity), and outliers are detected and rectified via tools like box plots to mitigate training bias. Furthermore, data standardization is applied to resolve dimensional inconsistencies across indicators.

(2) Feature Analysis: Time series plots are generated to visually inspect seasonal and cyclical patterns, such as the winter-spring surge of influenza. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are calculated to elucidate data autocorrelation and uncover latent temporal dependencies. Epidemiological domain knowledge is then applied to analyze correlations between external factors and disease metricsłsuch as the impact of temperature fluctuations on intestinal infectious diseasesłproviding a rationale for model selection.

(3) Model Selection and Adaptation: Model selection is tailored to the specific characteristics of the data. The Seasonal ARIMA (SARIMA) model is preferred for data exhibiting clear seasonality and linearity. Conversely, deep learning architectures like Long Short-Term Memory (LSTM) networks are superior for capturing the dynamics of complex, non-linear transmission. For scenarios involving comprehensive multi-factor analysis, Transformer-based derivatives are employed to effectively process holidays and external covariates. Additionally, hybrid approaches combining traditional statistical models with machine learning ensembles may be explored.

(4) Model Training: The preprocessed dataset is partitioned into training and testing subsets, typically following a 7:3 or 8:2 ratio. The training set is utilized to fit the model, with parameters fine-tuned to maximize the fit to historical disease patterns.

(5) Model Evaluation and Validation: Predictive precision is quantified on the test set using metrics including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Beyond statistical metrics, the model's capacity to forecast critical epidemiological eventsłsuch as outbreak onsets and trend inflection pointsłis evaluated against practical control needs. Cross-validation is employed to assess generalization capabilities; suboptimal performance necessitates an iterative adjustment of the model or its hyperparameters.

(6) Optimization and Refinement: Optimization algorithms, such as grid search or genetic algorithms, are deployed to identify optimal hyperparameters. Recognizing the dynamic nature of disease evolution, the model undergoes periodic retraining with updated datasets. Further refinements may involve incorporating novel variables or enhancing model architecture to better capture the complexities of the transmission process.

(7)Application and Dynamic Monitoring: The optimized model is deployed to forecast future trends, offering decision support for resource allocation and vaccination strategy formulation. In operational settings, new data is continuously assimilated to monitor performance in real-time. Should prediction deviations widen or transmission patterns shift, the model is promptly calibrated to ensure the timeliness and accuracy of the forecasts.

5. Conclusion

In an era marked by escalating public health threats and the pervasive emergence of novel infectious diseases, time series forecasting has established itself as a cornerstone of modern disease surveillance architectures. Predicated on an analysis of data characteristics, this review scrutinizes the evolution and application of forecasting methodologies, ranging from traditional statistical models (e.g., ARIMA, SARIMA, Exponential Smoothing) to machine learning approaches and advanced deep learning architectures (e.g., LSTM, GRU, Transformers, and their variants).

Collectively, traditional statistical models, distinguished by their theoretical maturity and interpretability, remain vital tools, particularly for scenarios involving limited data. Machine learning methods represent a step forward, overcoming the constraints of linear assumptions and facilitating the synthesis of multivariate data to address complex determinants. Notably, deep learning models have driven a paradigm shift; by leveraging robust automatic feature extraction, they excel in modeling long-range dependencies, non-linear dynamics, and multi-scale temporal structures, thereby becoming the dominant trend in contemporary research.

Despite these advancements, the field faces persistent challenges, including inherent data noise, the disruptive impact of sudden events, intricate seasonality, the ”black box” issue of interpretability, obstacles in fusing multi-source heterogeneous data, and constraints regarding model transferability and generalization. Therefore, the trajectory of future research must focus on rigorous theoretical innovation, sophisticated algorithm design, and the seamless adaptation of these models to complex, real-world public health scenarios.

6. Conflict of Interest

The authors declare that there are no conflict of interests, we do not have any possible conflicts of interest.

Acknowledgments. This research was supported by the Fundamental Research Funds for the Liaoning Universities(No.LJ202410166011).

References

1. Zhao X. Strategies and effectiveness of public health management in responding to sudden infectious disease incidents. Chin J Antituberc. 2024;46(S02):1-3. [in Chinese]
2. Okamoto S, Shimmei K, Okamura T. 1167 Disease burden and universal health coverage monitoring. Int J Epidemiol. 2021;50(Suppl 1):dyab168.508.
3. Zou X, Zhao H, Wang Q, et al. Research on influenza prediction in Chongqing based on ARIMA model. J Chongqing Med Univ. 2023;48(12):1425-1429. [in Chinese]
4. Dai H, Zhou N, Ren X, et al. Epidemiological characteristics and trend prediction of influenza subtypes based on autoregressive integrated moving average model. Dis Surveill. 2022;37(10):1338-1345. [in Chinese]
5. Wang B, Feng H, Huang P, et al. Application value of random forest model in prediction of severe hand, foot and mouth disease. J Zhengzhou Univ (Med Sci). 2018;53(06):747-751. [in Chinese]
6. Xie Y, Zhang H, Yang S. Prediction model of drug-induced phospholipidosis based on support vector machine. Chem Res Appl. 2011;23(06):696-701. [in Chinese]
7. Zhou H, Dong A, Li H, et al. Application of CNN-LSTM model optimized by intelligent algorithm in prediction of hand, foot and mouth disease. Mod Prev Med. 2024;51(08):1364-1369, 1376. [in Chinese]
8. Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time Series Analysis: Forecasting and Control. 3rd ed. San Francisco: Holden-Day; 1978.
9. Cai X, Wan Q, Wu Y, et al. ARIMA model for predicting the incidence trend of hand, foot and mouth disease in Zhabei District, Shanghai. Pract Prev Med. 2012;19(03):381-384. [in Chinese]
10. Zheng L, Zhai Y. Application of time series analysis in predicting incidence of infectious diseases. Strait J Prev Med. 2021;27(03):99-101. [in Chinese]
11. Wu Y, Liu W, Liang Q, et al. Time series analysis and prediction of hepatitis B epidemic trend in Jiangsu Province. Jiangsu Prev Med. 2010;21(06):15-17. [in Chinese]
12. Zheng S, Liu L, Bai W, et al. Application of autoregressive integrated moving average model in time series analysis in foodborne disease prediction. J Baotou Med. 2022;46(02):12-14. [in Chinese]
13. Wang X, Cheng X, Fang S, et al. Application of exponential smoothing model in prediction of influenza-like illness. China Trop Med. 2011;11(08):938-939. [in Chinese]
14. Zhu Y, Zhao Q, Feng W, et al. Application of exponential smoothing method in predicting the incidence trend of hepatitis A in Shanghai. Chin J Health Stat. 2013;30(01):31-33, 36. [in Chinese]
15. Liu T, Yao M, Huang J, et al. Comparison of model fitting and prediction effects of mumps incidence in China. Chin Gen Pract. 2020;23(11):1338-1343. [in Chinese]
16. Tang G, Zhang F, Lu W, et al. Application of exponential smoothing method in measles incidence prediction. Pract Prev Med. 2018;25(06):757-759. [in Chinese]
17. Zhou X. Elimination of schistosomiasis in Dongting Lake area moves towards a new era. Chin J Schistosomiasis Control. 2020;32(03):219-221. [in Chinese]
18. Bharadi VA, Alegavi SS. Analysis of Long Short Term Memory (LSTM) Networks in the Stateful and Stateless Mode for COVID-19 Impact Prediction. In: Dash S, Pani SK, Abraham A, Liang Y, editors. Advanced Soft Computing Techniques in Data Science, IoT and Cloud Computing. Studies in Big Data, vol 89. Cham: Springer; 2021. p. 203-228.
19. Mukhtar H, Taufiq RM, Herwinanda I, Winarso D, Hayami R. Forecasting Covid-19 time series data using the Long Short-Term Memory (LSTM). Int J Adv Comput Sci Appl. 2022;13(10):131026.
20. Li Z, Chen R, Lu X, et al. LSTM model for readmission prediction of heart failure patients. Comput Appl Softw. 2023;40(06):343-349. [in Chinese]
21. Lee JM, Hauskrecht M. Recent context-aware LSTM for clinical event time-series prediction. In: Proceedings of the 17th Conference on Artificial Intelligence in Medicine (AIME 2019); 2019 Jun 26-29; Poznan, Poland. Cham: Springer; 2019. p. 13-23.
22. Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, Zhang W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2021;35(12):11106-11115.
23. Wu H, Xu J, Wang J, Long M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In: Advances in Neural Information Processing Systems (NeurIPS). 2021;34:22419-22430.
24. Dong J, Wu H, Wang Y, Zhang L, Wang J, Long M. Metadata matters for time series: Informative forecasting with transformers. arXiv preprint arXiv:2410.03806. 2024.
25. Zhou T, Ye X, Liu F, Lu H. Lightweight directional transformer for X-ray-aided pneumonia diagnosis. Acta Opt Sin. 2023;43(13):1310001.
26. Wu N, Green B, Ben X, O'Banion S. Deep transformer models for time series forecasting: The influenza prevalence case. arXiv preprint arXiv:2001.08317. 2020.
27. Karagoz MA, Breton MD, El Fathi A. A comparative study of transformer-based models for multi-horizon blood glucose prediction. IFAC-PapersOnLine. 2025;59(2):155-160.