

Dynamic Prediction of Vaccination Rates and Regional Differences Based on LSTM Time Series Model

Min Song¹, Hua Li¹, Ling Yang²

¹Immunization Planning Department, Changzhi Disease Prevention and Control Center, Changzhi 102206, Shanxi, China

²Vaccination Department, Sudian Health Center, Changzhi 102206, Shanxi, China

Corresponding author: Min Song (e-mail: Min Song, 18235501119@163.com)

ABSTRACT Modeling spatiotemporal propagation and dynamic regional interactions is essential for intelligent decision-making in large-scale interconnected systems and communication-aware engineering applications. To overcome the limitations of conventional vaccination prediction methods that neglect dynamic spatial dependencies, this study proposes an LSTM-based forecasting framework integrating spatiotemporal graph convolution, dynamic spatial attention, and spatial regularization mechanisms. A dynamic spatial weight matrix combining geographic adjacency and population mobility is first constructed to characterize inter-regional coupling, after which graph convolution and adaptive attention are employed to capture time-varying spillover effects and evolving influence patterns. Experiments conducted on weekly vaccination data from 90 districts and counties in Zhejiang Province demonstrate that the proposed model reduces MAE and RMSE by 19.0% and 17.2%, respectively, compared with the conventional LSTM while significantly improving the spatial consistency of prediction residuals. Furthermore, the framework reveals hierarchical propagation characteristics from core urban clusters to transportation hubs and peripheral regions through interpretable dynamic attention and spillover analysis. By explicitly modeling spatial interaction and propagation mechanisms rather than isolated temporal sequences, the proposed approach provides a generalized methodology for spatiotemporal signal evolution analysis, network-aware prediction, and adaptive information propagation in complex distributed systems.

INDEX TERMS spatiotemporal propagation, graph convolutional LSTM, dynamic spatial attention, network-aware prediction, spatial dependence

I. INTRODUCTION

VACCINATION coverage rates serve as a fundamental metric for assessing the effectiveness of public health interventions [1-2]. Furthermore, vaccinations are a key component towards herd immunity. Many agree that observed vaccination behaviors vary by location due to the phenomenon known as “spatial dependence”. Vaccination behaviors exhibit significant spatial heterogeneity due to ‘spatial dependence,’ a phenomenon where local social interactions and population mobility (immigration/emigration) cause vaccination progress, policy implementation, and information dissemination in one region to influence its neighbors [3-4]. Most current predictions for vaccination behaviours are based on a region’s ability to function independently or only to include static geographic attributes; therefore, current approaches fail to accurately reflect the dynamic transfer of vaccination behaviours between regions, as such, having a negative impact on both the predictive capability of obtaining operation resources and the availability to accurately identify the factors that contribute to vaccine inequality across regions. There is substantial

theoretical value and pragmatic value to constructing a prediction framework that has deep spatio-temporal integrations of each regional difference, if it is to be used to optimize the allocation of vaccination resources, develop targeted intervention models. In the field of vaccination rate prediction and regional difference analysis, scholars have conducted a series of studies. Based on readily available indicators at the regional level, Carrieri et al. used machine learning models (especially the random forest algorithm) to effectively predict high-risk communities for vaccine hesitation, thereby helping policymakers implement precise regional vaccine advocacy interventions [5]. Wheeler et al. identified geospatial patterns and regional factors associated with HPV (Human Papilloma Virus) vaccination in specific geographic areas [6]. Dayaratna et al. used hierarchical Bayesian spatiotemporal modeling for spatial analysis, which is crucial for understanding the factors influencing disease transmission before the advent of vaccines [7]. Jentsch et al. constructed a socio-epidemiological coupling model of SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) transmission, suggesting that using

SARS-CoV-2 vaccines to block virus transmission may be more effective in preventing death than prioritizing vaccination of vulnerable age groups [8]. Although the above studies have made progress in verifying spatial effects, they generally have two limitations: the predictive model and spatial analysis are disconnected, and the spatial spillover intensity is treated as a static parameter, making it difficult to capture its dynamic evolution process. LSTM models have been gradually introduced into the field of vaccine vaccination rate prediction due to their advantages in time-series modeling. Altarawneh et al. proposed an interpretable convolutional neural network-long short-term memory (CNN-LSTM) model to predict the COVID-19 vaccination rate [9]. He et al. proposed an optimized long short-term memory (LSTM+) model to predict the cumulative number of confirmed COVID-19 cases in Germany, the United Kingdom, Italy and Japan [10]. Kumar et al. proposed an interpretable COVID-Twitter-BERT (A natural language processing model to analyze COVID-19 content on Twitter) and LSTM model to determine people's views on COVID-19 vaccination [11]. Burukanli et al. proposed a robust HyperMixer and long short-term memory network (LSTM) attention mechanism model HyperAttCov for COVID-19 virus mutation prediction [12]. In summary, while existing LSTM-based research has achieved breakthroughs in temporal prediction accuracy, it still has significant shortcomings in explicit modeling of spatial dependencies, dynamic capture of spatial spillover, and analysis of regional difference transmission mechanisms.

To address the research gaps in existing studies that fail to embed spatial dependency mechanisms into the temporal prediction framework and struggle to capture the dynamic evolution of spatial spillover and the transmission paths of regional differences, this paper proposes an LSTM prediction framework that integrates spatial dependency mechanisms. In this work, three key contributions are proposed. First, a spatiotemporal graph convolutional LSTM module is presented, which combines graph convolutional network and LSTM to provide a more accurate representation of the surrounding location's information during the time series modeling process than what is possible with traditional LSTMs that treat the regional data independently from one another. Second, a real-time dynamic spatial attention mechanism is proposed, which outputs spatial attention values as a function of the temporal correlation between the hidden states of each region. This mechanism allows to track how the influence of each location changes over time due to the presence of external influences. Third, a spatial regularization loss function is introduced, which is defined to reduce the difference in predicted values between adjacent locations and promote spatial consistency within the output of our models. The effectiveness of the proposal through experimental results collected from a dataset containing the vaccine dosage rates from 90 districts and counties in Zhejiang Province, China, is demonstrated, and the key points at which prediction errors propagate between adjacent

districts or counties are also quantified. The results provide a novel, understandable spatiotemporal fusion framework for predicting vaccine inoculation rates that can be used to support evidence-based public health decisions.

II. METHODS

A. OVERALL MODEL FRAMEWORK

To systematically present the overall architecture and data flow path of the LSTM prediction framework with fused spatial dependency mechanisms proposed in this paper, three core improved modules—spatiotemporal graph convolutional LSTM, dynamic spatial attention, and spatial regularization loss—are placed within the same logical flow, as shown in Figure 1.

As shown in Figure 1, the architecture consists of four core stages. First, the input and spatial construction layer lays the data foundation by integrating weekly vaccination rates, multidimensional covariates, and a dynamic spatial weight matrix A constructed based on geographical adjacency and population flow. Next, the spatiotemporal graph convolutional LSTM module utilizes graph convolutional networks (GCNs) to explicitly aggregate spatial context information of neighboring regions at each time step and input it into the LSTM unit to synchronously model temporal dependencies. Subsequently, a dynamic spatial attention mechanism calculates the temporal similarity of the hidden states of regions and evolves the spatial influence weight α in real time, thereby capturing the dynamic transmission patterns of spatial spillover effects. Finally, the system is optimized through a joint loss function integrating spatial regularization terms, improving prediction accuracy while constraining prediction differences between neighboring regions, ensuring the smoothness and consistency of the output results in spatial distribution.

B. PROBLEM DEFINITION AND SYMBOL SYSTEM

1) Regional and Temporal Definitions

Let the study area set be $R = \{1, 2, \dots, N\}$ ($N = 90$ in this paper), and the time window be $T = \{1, 2, \dots, T\}$ (weekly data, $T = 156$).

2) Features and Prediction Objectives

$X_t \in R^{N \times F}$ is defined as the feature matrix (vaccination rate and covariates) at time t , and $Y_t \in R^N$ as the true value vector of the vaccination rate at the corresponding time. Rolling prediction is adopted: using L historical time points ($L = 12$) to predict $\hat{Y}_{t+H}$ in the next H steps ($H = 1$ or 4).

3) Spatial Adjacency Matrix

A spatial adjacency matrix $A \in R^{N \times N}$ is constructed, where the elements a_{ij} are weighted based on a combination of geographical adjacency and population flow. After normalization, $\hat{A} = D^{-1/2} A D^{-1/2}$ is obtained, where D is a degree diagonal matrix.

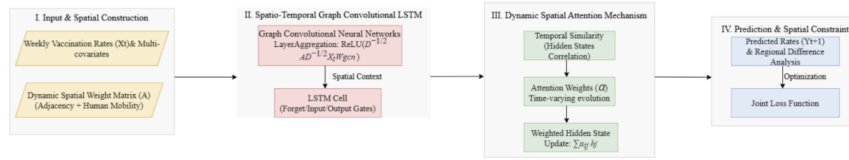


FIGURE 1. Overall model framework

4) Model Objectives and Evaluation Indicators

The nonlinear mapping $f: \{X_{t-L+1}, \dots, X_t\}, A \mapsto \hat{Y}_{t+H}$ is learned. Evaluation indicators include MAE, RMSE, MAPE (Mean Absolute Percentage Error), R^2 and the global Moran's index (Moran's I) of the prediction residuals [13].

C. SPATIAL RELATIONSHIP CONSTRUCTION

1) Spatial Weight Matrix Construction

The spatial adjacency matrix A is constructed using a composite weight of “geographical adjacency (Queen rule) + population flow”: $a_{ij} = b_{ij} \cdot m_{ij}$. After row normalization, it is symmetrically normalized to $\hat{A} = D^{-1/2} A D^{-1/2}$. In order to intuitively present the spatial adjacency matrix structure based on geographic adjacency and population flow [14] and verify the rationality of the setting of spatial dependence between regions, Figure 2 shows the top 50 districts and counties in Zhejiang Province. The connection within the urban agglomeration is strong, while the connection in the mountainous area is weak.

In Figure 2, the dark blue area with indices 0–25 in the upper left corner represents “Urban Agglomerations,” characterized by frequent geographical proximity and population movement, resulting in the strongest spatial spillover effect. The light yellow area with indices 25–50 in the lower right corner corresponds to “Mountainous Areas,” where spatial connections are relatively sparse due to geographical barriers. A dark diagonal line passing through the center represents the temporal correlation of the districts and counties themselves. The overall matrix provides core data support for the subsequent explicit modeling of spatial dependency mechanisms and analysis of the dynamic transmission of regional differences using the spatiotemporal graph convolutional LSTM model.

2) Temporal Similarity

This paper quantifies the morphological similarity of vaccination rate time series in various districts and counties of Zhejiang Province using DTW (Dynamic Time Warping) distance [15–16]. Using Ward link distance (the smaller the distance, the higher the “similarity” between samples), three clusters are identified: a rapid pattern in the northern plains, a transitional pattern in the central region, and a slow pattern in the southwestern mountainous areas. This confirms significant regional differentiation in the evolution trajectory of vaccination rates. Furthermore, by mapping the

cluster labels to a map of Zhejiang Province using a geographic distribution map—the rapid pattern is concentrated in urban clusters, while the slow pattern is distributed in mountainous areas—the clustering results exhibit obvious spatial clustering characteristics, as shown in Figure 3.

Figure 3(a) divides the time series of vaccination rates in 90 districts and counties of Zhejiang Province into three clusters using Ward link distance as the vertical axis: the blue cluster's link merging span is mainly concentrated below 0.5, representing the rapid vaccination pattern in the northern plains; the yellow cluster's link merging distance (0.5–1.5) is denser than other clusters, corresponding to the stable growth pattern in the central transition zone; the red cluster has the largest link distance span (approaching 2.5), indicating strong individual differences and heterogeneity in the vaccination trajectory of the southwestern mountainous areas. The geographical distribution map in Figure 3(b) maps the cluster labels to the map of Zhejiang Province. The blue clusters are concentrated in the plain urban clusters of Hangzhou-Jiaxing-Huzhou and Ningbo-Shaoxing; the red clusters are mainly located in the southwestern mountainous areas such as Lishui and Quzhou; the yellow clusters are distributed in the transitional zones such as the Jinhua-Quzhou Basin, showing significant spatial clustering characteristics. This verifies the premise that geographically adjacent areas have similar inoculation rate evolution trajectories and provides empirical evidence for the subsequent introduction of spatial dependence mechanisms.

D. SPATIOTEMPORAL GRAPH CONVOLUTIONAL LSTM MODULE

The core function of the spatiotemporal graph convolutional LSTM module is to explicitly aggregate the spatial information of neighboring regions at each time step, enabling the LSTM unit to perceive the spatial context between regions, thereby deeply integrating spatial dependence with temporal evolution [17–18]. The specific implementation consists of two steps: first, a graph convolution operation is performed on the feature matrix $X_t \in R^{N \times F}$ at the current time step. The symmetric normalized adjacency matrix $\hat{A} = D^{-1/2} A D^{-1/2}$ is used to aggregate the features of neighboring regions, and the learnable weight matrix $W_{gcn} \in R^{F \times H}$ maps the features to the hidden dimension H . Then, the spatially enhanced features are obtained through the ReLU (Rectified Linear Unit) activation function:

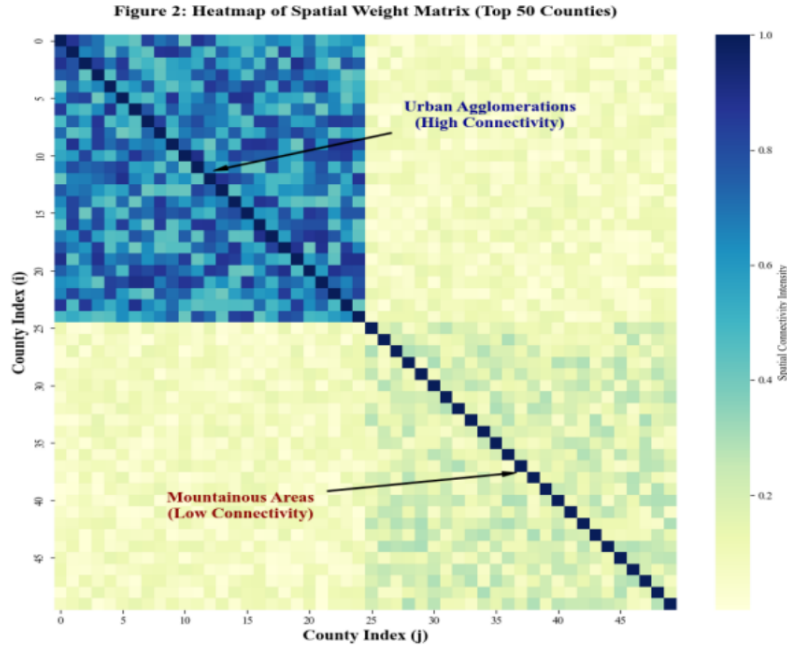


FIGURE 2. Heatmap of spatial weight matrix of the top 50 districts and counties in Zhejiang Province

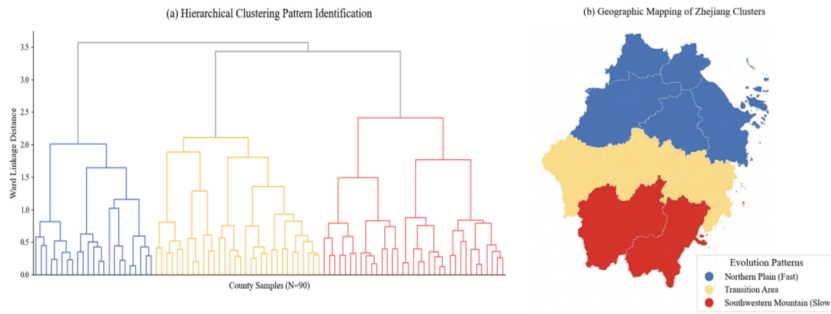


FIGURE 3. Hierarchical clustering results and spatial distribution of vaccination rates based on DTW distance

$$\tilde{X}_t = \text{ReLU}(\hat{A}X_t W_{gen}) \quad (1)$$

In Formula (1), $\tilde{X}_t \in R^{N \times H}$ integrates the information of each region itself and its neighbors. $\tilde{X}_t$ is the input of the LSTM unit, and the hidden state is updated according to the standard gating mechanism.

E. DYNAMIC SPATIAL ATTENTION MECHANISM

The dynamic spatial attention mechanism aims to capture the time-varying characteristics of spatial spillover intensity between regions, overcoming the limitation of traditional static spatial weight matrices in failing to reflect the evolutionary patterns of regional influence relationships during external shocks such as policy events and public opinion changes [19-20]. The core idea of this mechanism is: at each time step, attention weights are dynamically calculated based on the temporal similarity of the hidden states of the regions, and the hidden states are weighted and aggregated accordingly.

Specific implementation is as follows: for any two regions i and j , the attention score at time t is calculated based on the hidden states $h_{i,t-1}$ and $h_{j,t-1}$ of the previous time step. First, the two are concatenated, mapped to a scalar via the learnable vector $a \in R^{2H}$, and nonlinearity is introduced through the LeakyReLU activation function. Then, the attention weights are obtained by Softmax normalization [21-22]:

$$\alpha_{ij}^t = \frac{\exp(\text{LeakyReLU}(a^T[h_{i,t-1} \| h_{j,t-1}]))}{\sum_{k \in N(i)} \exp(\text{LeakyReLU}(a^T[h_{i,t-1} \| h_{k,t-1}]))} \quad (2)$$

In Formula (2), $N(i)$ represents the spatial neighborhood of region i (defined by the spatial adjacency matrix A), and $\|$ represents vector concatenation.

The attention weights α_{ij}^t reflect the influence strength of region j on region i at time t , and satisfy $\sum_{j \in N(i)} \alpha_{ij}^t = 1$. Subsequently, the hidden states of neighboring regions are

weighted and aggregated using this weight to obtain the enhanced region representation:

$$\hat{h}_{i,t} = \sum_{j \in N(i)} \alpha_{ij}^t \cdot h_{j,t-1} \quad (3)$$

In Formula (3), $h_{j,t-1}$ represents the attention score of any region j at time t based on the hidden state of the previous time step. This aggregation process allows the hidden state of each region to dynamically absorb information from its neighbors, and the influence weights are adaptively adjusted according to changes in external conditions.

F. SPATIAL REGULARIZATION LOSS FUNCTION

The core function of the spatial regularization loss function is to apply a smoothness constraint to the predicted values of adjacent regions while optimizing the prediction error, so that the model output has geographical consistency in space and avoids the unreasonable phenomenon of “geographically adjacent but abrupt prediction results” [23-24]. This design is based on the first law of geography (nearby things are more similar) and injects spatial autocorrelation as prior knowledge into the model training process.

The loss function consists of two parts: the prediction error term and the spatial regularization term. The prediction error term uses the mean squared error (MSE) to measure the difference between the model's predicted value $\hat{y}_{i,t}$ and the true value $y_{i,t}$:

$$L_{\text{pred}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (4)$$

In Formula (4), $(y_i - \hat{y}_i)^2$ is the square of the prediction bias. The spatial regularization term penalizes the prediction differences between adjacent regions and is defined as:

$$L_{\text{spatial}} = \sum_t \sum_{i=1}^N \sum_{j \in N(i)} w_{ij} (\hat{y}_{i,t} - \hat{y}_{j,t})^2 \quad (5)$$

In Formula (5), $N(i)$ represents the spatial neighborhood of region i , and w_{ij} is the spatial weight (consistent with the elements in the spatial adjacency matrix A). This formula is a weighted sum of the prediction differences between adjacent regions at all time steps. The total loss function is a weighted combination of the two terms:

$$L = L_{\text{pred}} + \lambda \cdot L_{\text{spatial}} \quad (6)$$

In Formula (6), the hyperparameter $\lambda > 0$ controls the strength of spatial smoothness. When λ increases, the model tends to produce spatially smooth prediction results; when $\lambda = 0$, the model degenerates into a standard LSTM (without spatial constraints).

To visually verify the effect of spatial regularization, Figure 4 selects a typical neighborhood containing three adjacent districts and compares the changes in the model prediction curves with and without spatial regularization.

In Figure 4(a), the neighboring region prediction curves of the standard LSTM (without spatial regularization) show obvious intersections and separations (e.g., County B deviates significantly from the true value and is inconsistent with its neighbors in weeks 15–20), violating the spatial intuition that geographically adjacent regions evolve similarly. In Figure 4(b), the prediction curve of the model in this paper (with spatial regularization) is closer to the true value and has a smooth transition without unreasonable abrupt changes. The comparison shows that the spatial regularization term effectively constrains the prediction differences between neighboring regions, ensures the spatial consistency of the output, and verifies its key role in improving the spatial rationality of the prediction.

III. EXPERIMENTS

A. DATASET DESCRIPTION

The research data covers 90 county-level administrative regions (including municipal districts, county-level cities, and counties) in Zhejiang Province, spanning from week 1 of 2021 to week 52 of 2023 (a total of 156 time points). Vaccination rate data is obtained from the weekly reporting system of the Zhejiang Provincial Center for Disease Control and Prevention, defined as the core variable: the ratio of the number of people who complete full vaccination each week to the resident population in each district/county. Covariates include: GDP per capita, number of doctors per thousand people, policy dummy variables (based on provincial health commission announcements marking events such as “strengthening vaccination promotion” and “community mobilization”), public opinion index, holiday dummy variables (statutory holidays and the three days before and after), and average weekly temperature. Based on the hierarchical clustering results mentioned above (Figure 3), the 90 districts/counties are divided into three regions: the rapid vaccination area in the northern plains, the transitional area in the central region, and the slow vaccination area in the southwestern mountainous region. The vaccination rates for each district/county are statistically analyzed, as shown in Table 1.

The Northern Plain Rapid Development Zone (32 districts/counties) has the highest average vaccination rate, 78.4%, as well as the least amount of internal variability (coefficient of variation (CV) 0.079). This reflects the fact that the Northern Plains region has made the most rapid and balanced progress with respect to vaccinations. The Central Transition Zone (28 districts/counties) has an average vaccination rate of 65.1% and a CV of 0.131. As a result, this area has moderate levels of vaccination progress compared to both the Northern Plains and the Southwestern Mountainous Slow Development Zone (30 districts/counties), which has

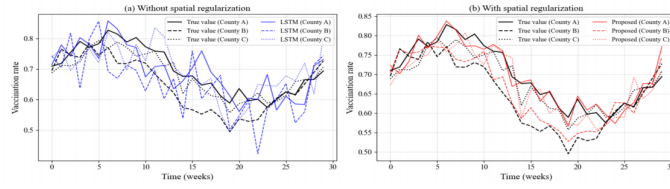


FIGURE 4. Comparison of prediction curves with and without spatial regularization

TABLE 1. Descriptive statistics of vaccination rates in each district/county

Region Type	Number of counties	Mean (%)	Standard deviation (%)	Coefficient of variation
Northern Plain (Fast)	32	78.4	6.2	0.079
Transition Area	28	65.1	8.5	0.131
Southwestern Mountain (Slow)	30	52.3	11.8	0.226
Overall (90 districts and counties)	90	65.7	13.4	0.204

Note: The mean and standard deviation are the average of the vaccination rates of each district/county over 156 weeks; the coefficient of variation = standard deviation /mean, reflecting the dispersion of vaccination rates within a region.

an average vaccination rate of only 52.3% and a CV of 0.226. Data shows that vaccinations remain low and relatively imbalanced within this region. Overall, the data from all three types of regions shows significant gradients in vaccination levels and internal balance within the same region; therefore, the results provide a very positive basis for training future models and analyzing disparities among geographical regions.

B. EXPERIMENTAL SETUP

1) Training/Validation/Test Set Division

To evaluate model performance, the dataset is divided into training, validation, and test sets in chronological order [25], with proportions of 70%, 15%, and 15%, respectively. Specifically, the first 109 weeks (week 1 of 2021 to week 5 of 2023) are used for training, the middle 24 weeks (week 6 to week 29 of 2023) for validation, and the last 23 weeks (week 30 to week 52 of 2023) for testing.

2) Evaluation Metrics

The model performance is evaluated using the following five metrics.

Let the test set contain M samples, y_i be the true value, $\hat{y}_i$ be the predicted value, $\bar{y}$ be the mean of the true values, N be the total number of regions, w_{ij} be the spatial weight, and $S_0 = \sum_{i=1}^N \sum_{j=1}^N w_{ij}$.

(1) Mean Absolute Error (MAE)

$$MAE = \frac{1}{M} \sum_{i=1}^M |y_i - \hat{y}_i| \quad (7)$$

(2) Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M (y_i - \hat{y}_i)^2} \quad (8)$$

(3) Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{M} \sum_{i=1}^M \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (9)$$

(4) Coefficient of determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^M (y_i - \hat{y}_i)^2}{\sum_{i=1}^M (y_i - \bar{y})^2} \quad (10)$$

(5) Global Moran's I . It is used to quantify the spatial autocorrelation of the predicted residual $\epsilon_i = y_i - \hat{y}_i$:

$$I = \frac{N}{S_0} \cdot \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (\epsilon_i - \bar{\epsilon})(\epsilon_j - \bar{\epsilon})}{\sum_{i=1}^N (\epsilon_i - \bar{\epsilon})^2} \quad (11)$$

In Formula (11), $\bar{\epsilon}$ is the mean of the residuals, and I takes values ranging from $[-1, 1]$. The larger the positive value, the stronger the spatial clustering of the residuals, indicating that the model has a better ability to capture spatial dependence.

C. COMPARISON MODELS

(1) ARIMA (AutoRegressive Integrated Moving Average model): A classic autoregressive moving average model that models each region independently and uses grid search to determine the optimal parameter (p, d, q) [26].

(2) Traditional LSTM: Standard LSTM network, which models each region independently and does not consider spatial dependencies between regions.

(3) LSTM + Static Spatial Weights: A fixed spatial adjacency matrix based on geographic adjacency is added to the LSTM input layer to achieve shallow fusion of spatial information.

(4) ST-GCN (Spatial-Temporal Graph Convolutional Networks): A representative spatiotemporal model that utilizes graph convolutional layers to capture fixed spatial dependencies and 1D temporal convolutional layers to extract temporal features [27]. Unlike the proposed model, ST-GCN

typically relies on a static graph structure and lacks a gated recurrent mechanism (like LSTM) or a dynamic attention layer to model time-varying spatial influences.

(5) The complete model in this paper: fused spatiotemporal graph convolutional LSTM, dynamic spatial attention and spatial regularization loss.

All models are implemented based on the PyTorch framework, using the same training/validation/test partitioning and hyperparameter settings (LSTM-type models have a hidden layer dimension of 64, and ARIMA parameters are optimized by grid search). To ensure robust results, each experiment is repeated 5 times. The results are reported as mean $\pm$ standard deviation, and paired t-tests are conducted to evaluate the performance differences between the proposed model and the baseline LSTM. A p-value < 0.05 is considered statistically significant.

D. ABLATION EXPERIMENT DESIGN

To verify the effectiveness of each core module of the proposed model, four ablation experiments are designed to compare the prediction performance and spatial consistency of different variants under the same dataset and hyperparameter settings. The complete model in this paper: fusing spatiotemporal graph convolutional LSTM, dynamic spatial attention mechanism, and spatial regularization loss, serves as a benchmark. Removal of the graph convolution module: The graph convolutional layer in the spatiotemporal graph convolutional LSTM is removed, and only the original feature X_t is directly input into the LSTM unit, while retaining dynamic attention and spatial regularization loss. This variant is used to verify the contribution of graph convolution to aggregating neighboring region information. Removal of dynamic attention (Static Weight Baseline): The dynamic spatial attention mechanism is replaced with a static weighting scheme where influence weights are fixed based on the time-invariant spatial adjacency matrix A . Specifically, the aggregation is reformulated as $\hat{h}_{i,t} = \sum_{j \in N(i)} a_{ij} \cdot h_{j,t-1}$, where a_{ij} is a fixed spatial weight. This variant, which retains graph convolution and spatial regularization, is specifically designed to isolate and quantify the unique contribution of temporal dynamics in the attention mechanism compared to traditional fixed-weight spatial modeling. Removing Spatial

Regularization Loss: The total loss function is degenerated to include only the prediction error term L_{pred} , i.e., $\lambda = 0$, while preserving graph convolution and dynamic attention. This variant is used to verify the constraint effect of spatial regularization on output space consistency.

IV. RESULTS

A. COMPARISON OF MODEL PREDICTION PERFORMANCE

To systematically quantify and evaluate the comprehensive performance of the proposed model and the comparison methods in terms of prediction accuracy and spatial con-

sistency, and to provide core data support for subsequent conclusions, Table 2 is presented.

In Table 2, the complete model in this paper outperforms all the comparative models and ablation variants in MAE (0.034), RMSE (0.048), MAPE (7.1%) and R^2 (0.892), and the Moran index of the predicted residuals (0.51) is significantly higher than that of other models, indicating that it has both high prediction accuracy and strong spatial consistency. In the ablation experiments, removing the graph convolution module increases the MAE to 0.038, RMSE to 0.053, and the Moran's index to 0.30, confirming the crucial role of graph convolution in spatial information aggregation. While removing dynamic attention (MAE 0.036, RMSE 0.051, Moran's index 0.41) outperforms LSTM+static spatial weights (0.038, 0.053, 0.28), it is still significantly weaker than the complete model. This indicates that graph convolution and spatial regularization provide basic spatial modeling capabilities, but the dynamic attention mechanism is indispensable for capturing the time-varying characteristics of spatial overflow. Removing spatial regularization reduces the Moran's index to 0.32, verifying the constraint effect of this loss term on the smoothness of the output space. In summary, Table 2 not only demonstrates the superiority of the proposed model compared to existing methods but also quantifies the independent contributions of the three core modules through ablation comparison.

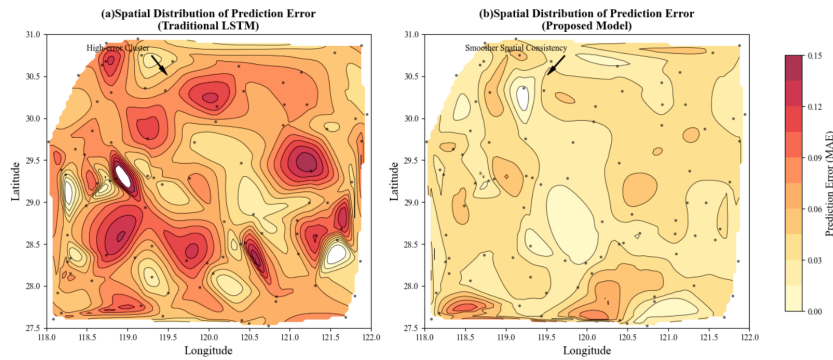
B. VERIFICATION OF IMPROVED SPATIAL CONSISTENCY

To verify the ability of the model in this paper to capture spatial dependence, the spatial distribution characteristics of the prediction residuals of the baseline LSTM and the model in this paper are compared. The global Moran index calculation results show that the Moran index of the prediction residuals of the baseline LSTM is 0.12, which is close to random distribution, indicating that its prediction error has no obvious spatial clustering; the Moran index of the prediction residuals of the model in this paper is 0.51, which is significantly higher than the baseline, indicating that the error shows strong spatial positive autocorrelation, that is, the prediction error of adjacent areas tends to be consistent, which is in line with the expectation of the first law of geography. However, a single index cannot reveal the spatial distribution pattern of the error. Figure 5 compares the spatial distribution of the prediction error of the baseline model and the model in this paper by mapping the error value to the geographic space and superimposing contour lines, which verifies the significant improvement in spatial consistency [28].

Figure 5(a) shows that the standard LSTM model, by treating each region as an independent entity, exhibits significant spatial dispersion and irregular abrupt changes in its prediction errors. Even between adjacent counties, significant clusters of high error values appear, contradicting the intuition that geographically adjacent regions should have similar evolutionary patterns. Figure 5(b) shows that after in-

TABLE 2. Comparison of model prediction performance

Model	MAE	RMSE	MAPE (%)	R^2	Moran's I
ARIMA	0.048	0.064	9.8	0.812	0.08
Traditional LSTM	0.042	0.058	8.7	0.853	0.12
LSTM + Static Spatial	0.038	0.053	7.9	0.871	0.28
ST-GCN	0.036	0.051	7.5	0.885	0.42
Proposed (Full)	0.034	0.048	7.1	0.892	0.51
w/o Graph Conv	0.038	0.053	7.9	0.870	0.30
w/o Dynamic Attention	0.036	0.051	7.5	0.884	0.41
w/o Spatial Reg	0.035	0.049	7.3	0.889	0.32

**FIGURE 5.** Spatial distribution comparison of prediction errors

corporating spatial regularization loss, the error distribution of the model in this paper not only has a lower overall value, but also exhibits a smooth transition in geographic space, proving that the model can better capture and constrain the dynamic spatial transmission process of vaccination rate while taking into account prediction accuracy.

C. DYNAMIC TRANSMISSION MECHANISM

Based on the hierarchical clustering results in Figure 3, the 90 districts and counties in Zhejiang Province are divided into three categories. To reveal the dynamic transmission mechanism of spatial spillover, this section analyzes it from three levels: predicted trajectory, spillover intensity, and attention evolution.

1) Predicted Trajectory and Regional Evolution Pattern

To visually present the differences in vaccination rate evolution paths among different regional types, verify the rationality of the hierarchical clustering results [29-30], and evaluate the model's fitting effect on each regional type, Figure 6 plots the vaccination rate curves and model predicted trajectories of three typical districts and counties, and superimposes confidence intervals, laying the regional classification foundation for subsequent spillover mechanism analysis.

Figure 6 shows the predicted trajectories of three typical districts and counties on the dataset. The vaccination rate in the core urban agglomeration rises rapidly and stabilizes at around 80%, with a narrow confidence interval, reflecting highly consistent vaccination behavior and high model fitting accuracy. Transportation hub areas show a steady growth trend with a moderate confidence interval.

Peripheral counties experience a slow start and greater fluctuations, with the widest confidence interval, reflecting the heterogeneity and uncertainty of vaccination behavior in these areas. The gradient differences in the predicted trajectories of the three types of regions intuitively verify the rationality of the regional classification and lay the foundation for subsequent spatial spillover analysis.

2) Regional Differentiation of Spatial Spillover Intensity

To quantitatively characterize the non-equilibrium of spatial spillover effects among different regional types and to examine the correlation between regional development level and spillover intensity [31], this paper defines two indicators: (1) Average spatial spillover coefficient: the average attention weight received by each district and county in the region from all neighboring areas (the attention weight α_{ij}^t of each week in the test set is averaged over time to obtain $\bar{\alpha}_{ij}$), i.e., $\frac{1}{n} \sum_{i \in \text{type}} \sum_{j \in N(i)} \bar{\alpha}_{ij}$, reflecting the overall intensity of the influence of the neighboring areas on this type of region; (2) Proportion of significant spillover neighboring areas: the proportion of the number of edges with an average attention weight $\bar{\alpha}_{ij} > 0.3$ between the neighboring areas and the total number of neighboring areas in this type of region, reflecting the proportion of strong spatial connections. Table 3 summarizes the average spatial spillover coefficient and the proportion of significant spillover neighboring areas for each regional type.

Table 3 shows that the core urban agglomeration has the highest average spatial spillover coefficient (0.28), with 73% of its neighboring areas exhibiting significant spillover, making it the main source of spillover. Transportation hub areas are second (0.18, 58%), while peripheral counties

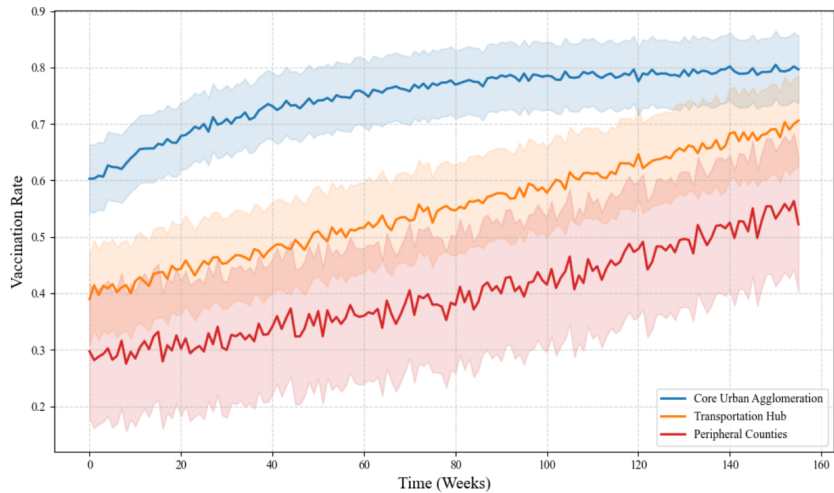


FIGURE 6. Clustering of predicted trajectories in different regions

TABLE 3. Quantification of spatial spillover effects of different regional types

Region Type	Average spatial spillover coefficient	Proportion of significant spillover neighbors (%)
Core urban agglomeration	0.28	73
Transportation hub	0.18	58
Peripheral counties	0.12	41

have the lowest (0.12, 41%). The results indicate a positive correlation between spatial spillover intensity and regional development level, with core urban agglomerations being the dominant spillover source and peripheral counties primarily acting as receivers, providing a quantitative basis for differentiated intervention.

3) Spatiotemporal Evolution of Dynamic Attention

Figure 7 provides a dynamic view of how the intensity of spatial spillover is impacted by time, along with academic research showing how an effective dynamic mechanism can be used to capture temporal relationships [32-33]. The heatmap displays three cross-sections across the city before and after the announcement of a policy (1 month prior; 3 months post-announcement). The heat map is complete from the accumulation of the spatial elements within the urban center to their radiation around them, as well as providing evidence of when these two elements reach equilibrium. Therefore, the abstract weights generated by the dynamic nature of attention can become represented as a visualized process over time, allowing for better clarity surrounding the nature of the transmission mechanism related to differences in regional areas.

Figures (a) and (b) show that before policy intervention, the matrix exhibits highly scattered and low-intensity characteristics, with no obvious weight clusters. In the map, the nodes in each region are lightly colored and uniformly sized, with almost no cross-regional connecting lines. Figures (c) and (d) show that one month after policy implementation, a distinct dark high-value area appears in the upper left corner of the matrix, indicating a rapid increase in mutual

influence between specific core regions. The corresponding nodes on the map become darker and larger. Figures (e) and (f) show that three months after the policy implementation, the high-weighted areas in the matrix further expanded and extended diagonally, reflecting the establishment of long-range dependencies. Several significant dark-colored high-influence centers also form on the map.

D. TRANSMISSION PATHS AND NETWORKS

To further reveal the micro-mechanisms of error propagation between regions, this section constructs a directed network based on dynamic attention weights and prediction errors [34-35], identifies influence centers and main transmission paths, and clarifies the specific transmission direction and hierarchical relationship of errors between regions, as shown in Figure 8.

Figure 8 clearly presents the hierarchical spillover characteristics of “core urban clusters—transportation hubs—peripheral counties”. The size of the nodes in the figure is proportional to their influence centrality. The core urban clusters, located at the center, act as the main source of error propagation, exhibiting extremely high out-degree characteristics. Through a high-intensity red transmission path (Primary Tier), they spread prediction bias to the transportation hubs in the central circle. Subsequently, these hub nodes further penetrate the influence to the peripheral counties through secondary paths (Secondary Tiers). This hierarchical network topology verifies that fluctuations in vaccination rates are not random diffusions, but rather follow a spatial evolution pattern of gradual transmission from the regional center to the periphery. While the dominance

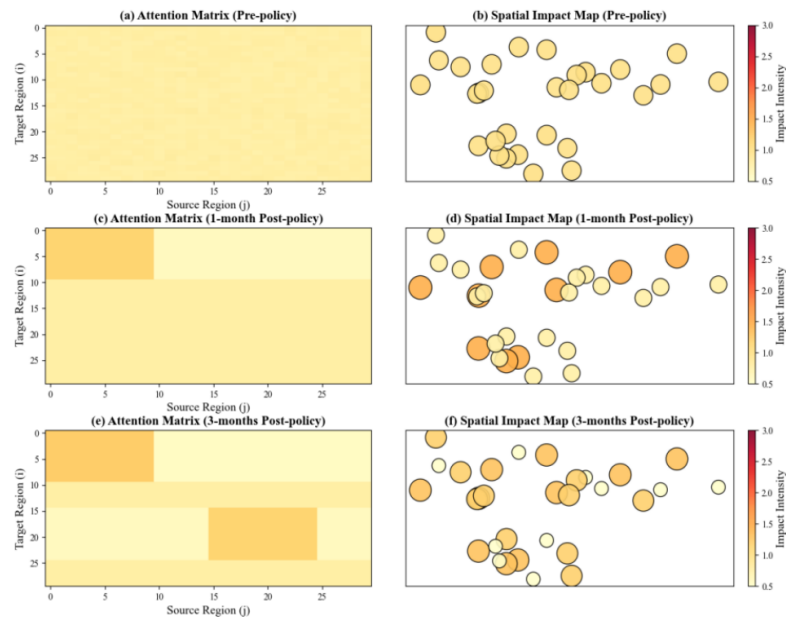


FIGURE 7. Spatiotemporal evolution of dynamic attention weights and spatial influence mapping

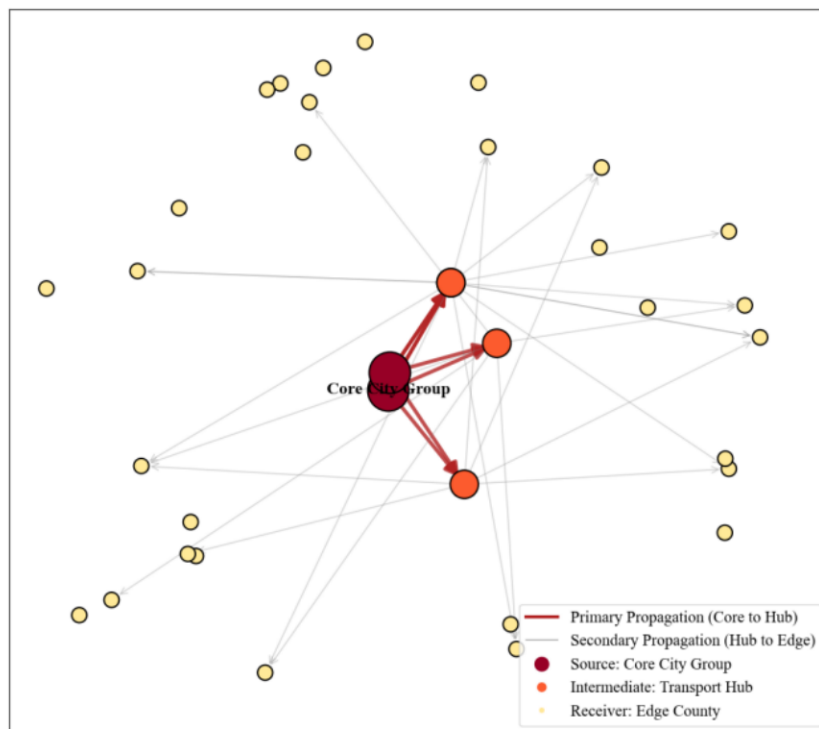


FIGURE 8. Directed network for spatial hierarchical error transmission based on dynamic attention weights

of major urban centers aligns with geographic intuition, the proposed model specifically quantifies the precise magnitude of these spillovers (e.g., Hangzhou's 0.52 vs. Shaoxing's 0.35), allowing for a prioritized hierarchy of intervention. Furthermore, the directed network in Figure 8 reveals that certain transportation hubs, despite lower total volume, serve as critical 'gatekeepers' or bridges that relay influence from

core clusters to remote peripheral counties, a structural insight that simple population-based intuition cannot provide.

To quantify and rank the influence centers in the error transmission network presented in Figure 8, and to provide clear key node targets for policy intervention, Table 4 reveals the core hub role of these regions in the error propagation network by calculating the weighted centrality of each

region.

TABLE 4. High-influence regions in Zhejiang Province and their spatial spillover intensity

Rank	Region Name	Spatial spillover coefficient
1	Hangzhou Urban Area	0.52
2	Ningbo Urban Area	0.47
3	Wenzhou Urban Area	0.43
4	Jinhua Urban Area	0.38
5	Shaoxing Urban Area	0.35

Note: Spatial spillover intensity is the sum of the average attention weights output by a region to all its neighbors.

Table 4 lists the top five high-influence regions in terms of spatial spillover intensity. Hangzhou city proper ranks first with a spillover intensity of 0.52, followed by secondary central cities such as Ningbo, Wenzhou, Jinhua, and Shaoxing. These regions are all core cities in Zhejiang Province with developed economies, dense populations, and prominent transportation hub status. Their high output weights in the dynamic attention network indicate that they are the main sources of spatial spillover and can have a significant radiating and driving effect on surrounding counties.

V. CONCLUSION

This paper constructs an LSTM prediction framework that integrates spatiotemporal graph convolution, dynamic spatial attention, and spatial regularization loss, focusing on dynamic prediction of vaccination rates and analysis of regional differences. The model is validated on data from 90 districts and counties in Zhejiang Province, achieving a 4.6% improvement in R2 compared to the baseline, and an increase in the Moran’s index of the predicted residual from 0.12 to 0.51, realizing a synergistic optimization of accuracy and spatial consistency. The research contributions include proposing an end-to-end spatiotemporal dependency fusion paradigm, revealing the gradient transmission pattern of spatial spillover intensity along the path from “core urban clusters → transportation hubs → peripheral counties,” and identifying high-influence nodes such as Hangzhou and Ningbo. By identifying high-influence nodes like Hangzhou and Ningbo, priority intervention targets are provided for the precise allocation of vaccine resources; the dynamic attention mechanism provides a visualized proxy for observing the shift in regional influence patterns before and after policy interventions, offering a qualitative reference for evaluating the reach of intervention measures; a resource allocation strategy that coordinates core supply and peripheral coverage is constructed based on the gradient transmission pattern from “core urban clusters → peripheral counties.” The limitations include the data being limited to a single province and not considering the fine-grained impact of inter-provincial migration and social media sentiment. Therefore, it can be extended to the fusion of multi-source data nationwide, introducing heterogeneous graph neural networks to handle multi-scale spatial relationships.

AUTHOR CONTRIBUTIONS

Conceptualization – Min Song, Hua Li and Ling Yang; methodology – Min Song, Hua Li and Ling Yang; investigation – Min Song, Hua Li and Ling Yang; writing–original draft preparation – Min Song, Hua Li and Ling Yang. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] J. Charron, A. Gautier, and C. Jestin, “Influence of information sources on vaccine hesitancy and practices,” *Medecine et maladies infectieuses*, vol. 50, no. 8, pp. 727-733, 2020, doi: 10.1016/j.medmal.2020.01.010.

[2] V. Lazarus J, C. Ratzan S, A. Palayew, et al., “A global survey of potential acceptance of a COVID-19 vaccine,” *Nature medicine*, vol. 27, no. 2, pp. 225-228, 2021, doi: 10.1038/s41591-020-1124-9.

[3] T. Das, “Unveiling spatial heterogeneity: a study of diverse child immunization indicators across Indian districts,” *Child Indicators Research*, vol. 17, no. 4, pp. 1615-1659, 2024, doi: 10.1007/s12187-024-10135-4.

[4] H. Chang M and T. Tassier, “Spatially heterogeneous vaccine coverage and externalities in a computational model of epidemics,” *Computational Economics*, vol. 58, no. 1, pp. 27-55, 2021, doi: 10.1007/s10614-019-09918-7.

[5] V. Carrieri, R. Lagravinese, and G. Resce, “Predicting vaccine hesitancy from area-level indicators: A machine learning approach,” *Health Economics*, vol. 30, no. 12, pp. 3248-3256, 2021, doi: 10.1002/hec.4430.

[6] C. Wheeler D, A. Miller C, K. Do E, et al., “Identifying area-level disparities in human papillomavirus vaccination coverage using geospatial analysis,” *Cancer Epidemiology, Biomarkers & Prevention*, vol. 30, no. 9, pp. 1689-1696, 2021, doi: 10.1158/1055-9965.EPI-21-0331.

[7] D. Dayaratna K, D. Gonszorowski, and M. Kolesar, “Hierarchical Bayesian spatio-temporal modeling of COVID-19 in the United States,” *Journal of Applied Statistics*, vol. 50, no. 11-12, pp. 2663-2680, 2023, doi: 10.1080/02664763.2022.2069232.

[8] C. Jentsch P, M. Anand, and T. Bauch C, “Prioritising COVID-19 vaccination in changing social and epidemiological landscapes: a mathematical modelling study,” *The Lancet Infectious Diseases*, vol. 21, no. 8, pp. 1097-1106, 2021, doi: 10.1016/S1473-3099(21)00057-8.

[9] L. Altarawneh, H. Wang, and Y. Jin, “COVID-19 vaccine prediction based on an interpretable CNN-LSTM model with three-stage feature engineering,” *Health and Technology*, vol. 14, no. 6, pp. 1241-1261, 2024, doi: 10.1007/s12553-023-00797-5.

[10] M. He, W. Zhu W, Z. Chen H, et al., “Application of optimized LSTM in prediction of the cumulative confirmed cases of COVID-19,” *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 27, no. 13, pp. 1893-1905, 2024, doi: 10.1080/10255842.2023.2264438.

[11] A. Kumar, P. Singh J, and K. Singh A, “Explainable BERT-LSTM stacking for sentiment analysis of COVID-19 vaccination,” *IEEE Transactions on Computational Social Systems*, vol. 12, no. 3, pp. 1296-1306, 2023, doi: 10.1109/TCSS.2023.3329664.

[12] M. Burukanli and N. Yumuşak, “COVID-19 virus mutation prediction with LSTM and attention mechanisms,” *The Computer Journal*, vol. 67, no. 10, pp. 2934-2944, 2024, doi: 10.1093/comjnl/bxae058.

[13] R. Westerholt, “A simulation study to explore inference about global Moran’s I with random spatial indexes,” *Geographical Analysis*, vol. 55, no. 4, pp. 621-650, 2023, doi: 10.1111/gean.12349.

[14] P. Phang, S. Aslam, J. Labadin, et al., “Spatial autocorrelation analysis of infectious disease incidence rates at state and district level using supra-adjacency weights matrix,” *Health*, vol. 13, no. 2, pp. 456-470, 2025, doi: 10.13189/ujph.2025.130217.

- [15] Z. Tao, Q. Xu, X. Liu, et al., "An integrated approach implementing sliding window and DTW distance for time series forecasting tasks," *Applied Intelligence*, vol. 53, no. 17, pp. 20614-20625, 2023, doi: 10.1007/s10489-023-04590-9.
- [16] C. Holder, M. Middlehurst, and A. Bagnall, "A review and evaluation of elastic distance functions for time series clustering," *Knowledge and Information Systems*, vol. 66, no. 2, pp. 765-809, 2024, doi: 10.1007/s10115-023-01952-0.
- [17] M. Kaselimi, N. Doulamis, A. Voulodimos, et al., "Context aware energy disaggregation using adaptive bidirectional LSTM models," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3054-3067, 2020, doi: 10.1109/TSG.2020.2974347.
- [18] T. Zhou, A. Tao, L. Sun, et al., "Behavior recognition based on the improved density clustering and context-guided Bi-LSTM model," *Multimedia Tools and Applications*, vol. 82, no. 29, pp. 45471-45488, 2023, doi: 10.1007/s11042-023-15501-y.
- [19] H. Guo M, X. Xu T, J. Liu J, et al., "Attention mechanisms in computer vision: A survey," *Computational visual media*, vol. 8, no. 3, pp. 331-368, 2022, doi: 10.1007/s41095-022-0271-y.
- [20] G. Brauwiers and F. Frasincar, "A general survey on attention mechanisms in deep learning," *IEEE transactions on knowledge and data engineering*, vol. 35, no. 4, pp. 3279-3298, 2021, doi: 10.1109/TKDE.2021.3126456.
- [21] D. Zou, Y. Cao, D. Zhou, et al., "Gradient descent optimizes over-parameterized deep ReLU networks," *Machine learning*, vol. 109, no. 3, pp. 467-492, 2020, doi: 10.1007/s10994-019-05839-6.
- [22] P. Blanchard, J. Higham D, and J. Higham N, "Accurately computing the log-sum-exp and softmax functions," *IMA Journal of Numerical Analysis*, vol. 41, no. 4, pp. 2311-2330, 2021, doi: 10.1093/imanum/draa038.
- [23] Q. Wang, Y. Ma, K. Zhao, et al., "A comprehensive survey of loss functions in machine learning," *Annals of Data Science*, vol. 9, no. 2, pp. 187-212, 2022, doi: 10.1007/s40745-020-00253-5.
- [24] R. Moradi, R. Berangi, and B. Minaei, "A survey of regularization strategies for deep models," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 3947-3986, 2020, doi: 10.1007/s10462-019-09784-7.
- [25] T. Burzykowski, M. Geubbelmans, A. J. Roussea, et al., "Validation of machine learning algorithms," *American journal of orthodontics and dentofacial orthopedics*, vol. 164, no. 2, pp. 295-297, 2023, doi: 10.1016/j.ajodo.2023.05.007.
- [26] K. Albeladi, B. Zafar, and A. Mueen, "Time series forecasting using LSTM and ARIMA," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 1, pp. 313-320, 2023, doi: 10.14569/ijacsa.2023.0140133.
- [27] M. Lovanshi and V. Tiwari, "Human skeleton pose and spatio-temporal feature-based activity recognition using ST-GCN," *Multimedia Tools and Applications*, vol. 83, no. 5, pp. 12705-12730, 2024, doi: 10.1007/s11042-023-16001-9.
- [28] H. Bai, H. Wang, D. Li, et al., "Information consistency-based measures for spatial stratified heterogeneity," *Annals of the American Association of Geographers*, vol. 113, no. 10, pp. 2512-2524, 2023, doi: 10.1080/24694452.2023.2223700.
- [29] X. Ran, Y. Xi, Y. Lu, et al., "Comprehensive survey on hierarchical clustering algorithms and the recent developments," *Artificial Intelligence Review*, vol. 56, no. 8, pp. 8219-8264, 2023, doi: 10.1007/s10462-022-10366-3.
- [30] Y. Komaru, T. Yoshida, Y. Hamasaki, et al., "Hierarchical clustering analysis for predicting 1-year mortality after starting hemodialysis," *Kidney international reports*, vol. 5, no. 8, pp. 1188-1195, 2020, doi: 10.15276/imms.v13.no1-2.5.
- [31] T. Shi, Y. Qiao, and Q. Zhou, "Spatiotemporal evolution and spatial relevance of urban resilience: Evidence from cities of China," *Growth and Change*, vol. 52, no. 4, pp. 2364-2390, 2021, doi: 10.1111/grow.12554.
- [32] Z. Wu, D. Zheng, S. Pan, et al., "TraverseNet: Unifying space and time in message passing for traffic forecasting," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 2, pp. 2003-2013, 2022, doi: 10.1109/tnnls.2022.3186103.
- [33] G. Spadon, S. Hong, B. Brandoli, et al., "Pay attention to evolution: Time series forecasting with deep graph-evolution learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 5368-5384, 2021, doi: 10.1109/TPAMI.2021.3076155.
- [34] X. Yuan, L. Li, Y. A. W. Shardt , et al., "Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 5, pp. 4404-4414, 2020, doi: 10.1109/TIE.2020.2984443.
- [35] S. Chaudhari, V. Mithal, G. Polatkan, et al., "An attentive survey of attention models," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 12, no. 5, pp. 1-32, 2021, doi: 10.1145/3465055.