ChatGPT-Assisted Deep Learning Models for Influenza-Like Illness Prediction in Mainland China

Wudi wei

Submitted to: Journal of Medical Internet Research
on: March 24, 2025

Table of Contents

ChatGPT-Assisted Deep Learning Models for Influenza-Like Illness Prediction in Mainland China

Wudi wei[1]

[1] Guangxi Medical University Nanning CN

Corresponding Author:
Wudi wei

Guangxi Medical University
22 Shuangyong Road
Nanning
CN

Abstract

Background: Background
While deep learning shows potential for influenza forecasting, implementation complexity persists. Large language models like ChatGPT enable automated code generation and model optimization, potentially lowering technical barriers in epidemiological research.

Objective: To evaluate the effectiveness of deep learning models in predicting Influenza-Like Illness (ILI) positive rates in Mainland China and explore the utility of ChatGPT as a development assistant in model construction and optimization.

Methods: ILI positive rate data spanning from 2014 to 2024 were obtained from the Chinese CDC database. Five deep learning architectures, Long Short-Term Memory (LSTM), Neural Basis Expansion Analysis for Time Series (N-BEATS), Transformer, Temporal Fusion Transformer (TFT), and Time-series Dense Encoder (TiDE), were implemented with ChatGPT's assistance for code generation, debugging, and optimization. Models were trained on data from 2014 to 2023 and evaluated on 2024 data (weeks 1-39) using Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) metrics.

Results: ILI cases in China exhibited clear seasonal patterns withwinter peaks and summer troughs, showing marked fluctuations during 2020-2022. TIDE demonstrated superior performance nationally (MAE = 5.551, MSE = 43.976, MAPE = 72.413%) and in southern China (MAE = 7.554, MSE = 89.708, MAPE = 74.475%). Northern region predictions were challenging across all models, with TIDE still performing best (MAE = 4.131, MSE = 28.922) despite high percentage errors (MAPE > 400%). ChatGPT significantly accelerated model development through automated code generation and optimization suggestions.

Conclusions: Deep learning models, particularly TIDE, show promise for ILI forecasting in China, with performance varying by region. Large language models like ChatGPT can substantially enhance research efficiency in epidemic prediction modeling, offering a scalable approach for public health preparedness.

(JMIR Preprints 24/03/2025:74423)
DOI: https://doi.org/10.2196/preprints.74423

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript

ChatGPT-Assisted Deep Learning Models for Influenza-Like Illness Prediction in Mainland China

Weihong Huang [#1] , Wudi Wei [#1,2] , Xiaotao He [1], Baili Zhan [1] , Xiaoting Xie [1] , Meng Zhang [1],Shiyi Lai [1] , Zongxiang Yuan [1] , Jingzhen Lai [1,2] , Rongfeng Chen [1,2] , Junjun Jiang [*1,2] , Li Ye [*1,2] , Hao Liang [*1,2]

1 Guangxi Key Laboratory of AIDS Prevention and Treatment & School of Public Health, Guangxi Medical University, Nanning, Guangxi, China
2 Joint Laboratory for Emerging Infectious Diseases in China (Guangxi)-ASEAN, Life Sciences Institute, Guangxi Medical University, Nanning, Guangxi, China

# These authors contributed equally to this work.
* Corresponding authors
jiangjunjun@gxmu.edu.cn(JJ);
yeli@gxmu.edu.cn(LY);
lianghao@gxmu.edu.cn(HL);

Abstract
Background
While deep learning shows potential for influenza forecasting, implementation complexity persists. Large language models like ChatGPT enable automated code generation and model optimization, potentially lowering technical barriers in epidemiological research.
Objective
To evaluate the effectiveness of deep learning models in predicting Influenza-Like Illness (ILI) positive rates in Mainland China and explore the utility of ChatGPT as a development assistant in model construction and optimization.
Methods
ILI positive rate data spanning from 2014 to 2024 were obtained from the Chinese CDC database. Five deep learning architectures, Long Short-Term Memory (LSTM), Neural Basis Expansion Analysis for Time Series (N-BEATS), Transformer, Temporal Fusion Transformer (TFT), and Time-series Dense Encoder (TiDE), were implemented with ChatGPT's assistance for code generation, debugging, and optimization. Models were trained on data from 2014 to 2023 and evaluated on 2024

data (weeks 1-39) using Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) metrics.

Results

ILI cases in China exhibited clear seasonal patterns withwinter peaks and summer troughs, showing marked fluctuations during 2020-2022. TIDE demonstrated superior performance nationally (MAE = 5.551, MSE = 43.976, MAPE = 72.413%) and in southern China (MAE = 7.554, MSE = 89.708, MAPE = 74.475%). Northern region predictions were challenging across all models, with TIDE still performing best (MAE = 4.131, MSE = 28.922) despite high percentage errors (MAPE > 400%). ChatGPT significantly accelerated model development through automated code generation and optimization suggestions.

Conclusion

Deep learning models, particularly TIDE, show promise for ILI forecasting in China, with performance varying by region. Large language models like ChatGPT can substantially enhance research efficiency in epidemic prediction modeling, offering a scalable approach for public health preparedness.

Key Words：Influenza-Like Illness; Deep learning; ChatGPT; Prediction; Seasonal pattern

Introduction

Influenza-Like Illness (ILI)(1) serves as a crucial indicator for monitoring influenza activity globally. It is characterized by fever (body temperature ≥ 38°C) along with symptoms such as cough or sore throat. However, these symptoms are not exclusive to influenza, as other respiratory infections can present similarly(2). The World Health Organization (WHO) estimates that seasonal influenza causes 290,000 - 650,000 respiratory deaths annually(3,4).Beyond mortality, seasonal influenza imposes substantial economic burdens through healthcare costs and productivity losses(5,6). Effective management of seasonal and pandemic influenza requires early detection through accurate surveillance and timely responses to prevent overcrowding in healthcare facilities(7).

Traditional surveillance systems often suffer from delays, limiting their ability to predict future epidemics. This highlights the need for advanced technologies to forecast ILI and influenza trends, enabling better early warning and intervention strategies, especially during pandemics(9,10). Deep learning models, such as Long Short-Term Memory (LSTM) (11), Neural Basis Expansion Analysis for Time Series (N-BEATS)(12), Transformer(13), Temporal Fusion Transformer (TFT)(14), and Time-series Dense Encoder (TIDE)(15), have shown promise in analyzing time series data for disease trend prediction. Although these models have demonstrated excellent performance in various applications, including epidemic prediction, their application in ILI prediction, particularly in real - time public health monitoring, remains limited(16). On one hand, the validity of these deep learning approaches for ILI forecasting has not yet been definitively established, as there remains insufficient comprehensive research and empirical evidence to thoroughly evaluate their predictive capabilities across diverse scenarios and geographical regions(17). On the other hand, implementing deep learning methodologies in predictive analytics presents a significant barrier, particularly for frontline health practitioners lacking programming expertise(18,19). These individuals often face overwhelming challenges in understanding and implementing complex deep learning algorithms, making widespread adoption of deep learning-based predictive systems-even when proven effective-extremely difficult. Nonetheless, the emergence of large language models (LLMs)(20) offers promising potential to substantially reduce these barriers.

Large language models like ChatGPT (21), have recently emerged as powerful tools across multiple domains, including scientific research, natural language processing, and epidemic prediction. Built on advanced machine learning architectures, they can support tasks such as code generation, data analysis, and model optimization(22). In epidemiological research, LLMs(23) can potentially democratize access to sophisticated analytical techniques by generating functional code, troubleshooting implementation issues, and suggesting optimizations without requiring extensive programming knowledge from researchers(24). Despite their current use in many fields, the application of large language models in epidemic prediction and early warning systems is still in its infancy. However, their development suggests they may become integral to future research, offering new paradigms for model development and analysis in public health.

In this study, we utilized LSTM, Transformer, N - BEATS, TFT, and TIDE models to predict the ILI positive rate in China. Our objectives were to identify the most accurate model for ILI trend forecasting and investigate the auxiliary role of ChatGPT in improving the research process, specifically in code generation, debugging, and model optimization. By leveraging ChatGPT for these tasks, we aimed to evaluate whether such AI - assisted approaches could offer a new and efficient paradigm for epidemic prediction, enhancing productivity and scalability in future research.

Methods

Ethical Approval

The data used in this study were sourced from a publicly accessible secondary database. Therefore, formal ethical approval was not required.

Data Source

The data on the positive rate of Influenza-Like Illness (ILI) cases from the southern and northern regions of mainland China, as well as nationwide, spanning from the 1st week of 2014 to the 39th week of 2024, were obtained from the public database of the Chinese National Influenza Center(25) (https://ivdc.chinacdc.cn/cnic/).

Data Preprocessing and Model Input

The dataset was divided into a training set and a validation set. Specifically, the training set comprised data from the 1st week of 2014 to the 52nd week of 2023, which was used to develop and train the models. The validation set covered data from the 1st week to the 39th week of 2024, used to evaluate the models' predictive performance on unseen, future data. This chronological partitioning permitted evaluation of the models' capacity to extrapolate from historical patterns to accurately forecast emerging ILI trends in real-world settings.

Construction of LSTM Models

Long Short-Term Memory (LSTM) networks, a specialized type of Recurrent Neural Network (RNN), were implemented to model temporal dependencies in ILI positive rate time series. The LSTM architecture employs memory cells with gating mechanisms that regulate information flow, enabling the model to capture both short-term fluctuations and long-term seasonal patterns critical for ILI forecasting. The core computational mechanism of our LSTM implementation can be expressed as:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$

where c_t represents the cell state at time t, f_t is the forget gate activation, i_t is the input gate activation, $\tilde{c}_t$ is the candidate cell state,and $\odot$ denotes element-wisemultication. This formulation allows the model to selectively retain relevant historical information while discarding noise, making it particularly suitable for capturing the complex temporal patterns characteristic of ILI epidemiological data. The LSTM networks were configured with bidirectional layers to process sequence information in both forward and backward directions, enhancing the model's ability to detect contextual patterns in the time series.

Construction of N-BEATS Models

The Neural Basis Expansion Analysis for Time Series (N-BEATS) model was implemented with a hierarchical architecture utilizing fully connected neural networks with dual-residual connections. This deep learning framework decomposes time series into interpretable components through iterative prediction and backcast residuals. The core prediction mechanism can be formulated as:

$$\hat{y}_{t+1:t+H} = \sum_{i=1}^{N} \theta_i^f (x_{t-L:t})$$

where $\hat{y}_{t+1:t+H}$ represents the H-step forecast, θ_i^f denotes the forward projection function of the i-th stack, and $x_{t-L:t}$ is the historical input sequence of length L. This architecture enables N-BEATS to simultaneously capture trend, seasonality, and irregular components of ILI time series without requiring explicit feature engineering. The model's stack-based design proved particularly effective for handling the non-stationary characteristics and complex seasonal patterns observed in regional ILI positive rates.

Construction of TFT Models

The Temporal Fusion Transformer (TFT) model was configured to leverage both recurrent networks and self-attention mechanisms for multi-horizon ILI forecasting. This architecture integrates variable selection networks, static covariate encoders, and multi-head attention components to process temporal relationships at different scales. The interpretable multi-head attention mechanism, central to the TFT architecture, can be expressed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q, K, and V represent query, key, and value matrices derived from the input sequence, and d_k is the dimension of the key vectors. This formulation allows the model to dynamically weight historical observations based on their relevance to future ILI predictions. The TFT implementation incorporated static metadata alongside temporal features, enabling it to account for regional differences in ILI transmission patterns while maintaining computational efficiency.

Construction of TIDE Models

The Time - series Dense Encoder (TIDE) model was implemented using an encoder-decoder framework composed of multiple multi-layer perceptrons (MLPs) optimized for long-range temporal dependencies. This architecture efficiently processes temporal information through dimensionality reduction and sequential transformation. The core projection mechanism in TIDE can be expressed as:

$$Z_t = \phi(W_z X_t + b_z)$$

where *Zt* represents the encoded state at time t, ϕ is a non-linear activation function, *Wz* is the projection weight matrix, and *Xt* is the input vector of ILI features. This formulation enables efficient parameter sharing across time steps while preserving temporal information critical for accurate ILI forecasting. The TIDE model's architectural simplicity provided computational advantages for long-term forecasting while maintaining robustness to the non-stationary characteristics of ILI time series data.

Construction of Transformer Models

The Transformer model was implemented using a self-attention-based encoder-decoder architecture optimized for sequential ILI data processing. Unlike recurrent networks, Transformers process entire sequences simultaneously through positional encodings and multi-head attention mechanisms. The scaled dot-product attention, fundamental to the Transformer's operation, can be represented as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where the softmax function normalizes attention weights across the sequence. For ILI forecasting, this mechanism enables the model to identify complex dependencies between observations at different time points regardless of their temporal distance.

Our implementation incorporated an encoder stack that processed historical ILI patterns and a decoder stack that generated multi-step forecasts, with masking strategies to prevent information leakage from future to past time steps.

ChatGPT-Assisted Model Development

In this study, we employed the ChatGPT large language model as an auxiliary tool to systematically support the development of deep learning models. ChatGPT's application primarily involved two key aspects:

(1) **Code Generation:** We utilized structured prompt templates to generate initial code frameworks for each model. Standardized query formats were established (e.g., "Generate a Python script for time-series prediction using the Darts library and the Transformer model, including complete steps for data loading, preprocessing, model definition, training, prediction, and evaluation") to ensure consistency and comprehensiveness in code generation. This methodology facilitated efficient implementation of complex model architectures without extensive manual coding.

(2) **Error Diagnosis:** A systematic error analysis workflow was implemented, wherein encountered code errors and their contextual information were submitted to ChatGPT for diagnostic assessment. This approach proved particularly valuable during data preprocessing stages involving time series transformations, where it helped identify and resolve various anomalies in datetime format conversion and data scaling processes. The error types and resolution strategies were documented and categorized to establish a standardized debugging protocol.

All interactions with ChatGPT adhered to predefined prompt templates to ensure methodological reproducibility. The complete prompt structures and exemplars are available in the supplementary materials. This approach provides a replicable AI-assisted analytical framework for epidemiological prediction research, potentially reducing technical barriers to deep learning applications in public health contexts.

Software and Libraries

All models were implemented in Python using TensorFlow and Keras for LSTM, N-BEATS, TFT, and Transformer models, with PyTorch used for some prototyping tasks. Time series forecasting and model evaluation were carried out using the Darts library. Data preprocessing and manipulation were done with NumPy and Pandas, while Matplotlib and Seaborn were used for visualization. SciPy was used for statistical analysis and hyperparameter optimization. Additionally, ChatGPT was utilized for code generation, debugging, and model optimization throughout the study.

Model Comparison

To evaluate the performance of the models, we used several metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Smaller values of these metrics indicate superior predictive accuracy. The formulas for these metrics are as follows:

$$\text{MSE}=\frac{1}{N}\sum_{i=1}^{N}\left(Y_i-\widehat{Y}_i\right)^2$$

$$\text{MAE}=\frac{1}{N}\sum_{i=1}^{N}\left|Y_i-\widehat{Y}_i\right|$$

$$\text{MAPE}=\frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{Y_i-\widehat{Y}_i}{Yi}\right|$$

where ˆyi represents the predictive value and yi represents the actual value.

Results

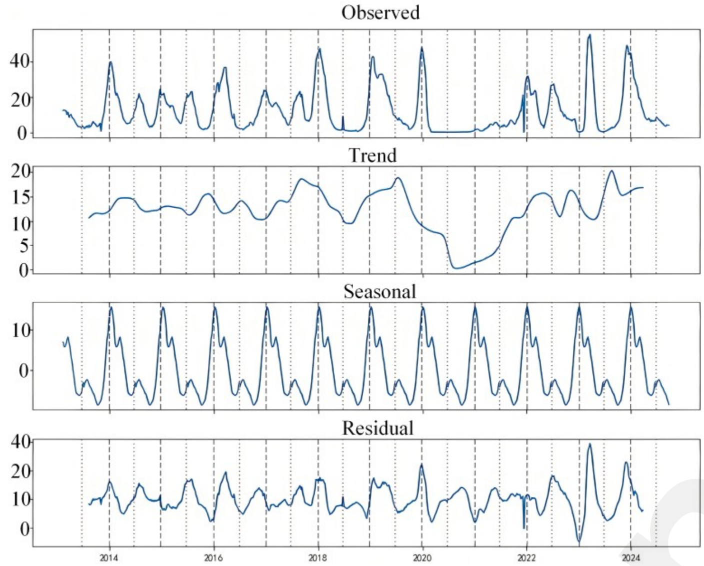
Characteristics of Influenza-Like Symptoms in Mainland China

We systematically collected influenza-like illness (ILI) case data from southern China, northern China, and mainland China as a whole between the 1st week of 2014 and the 39th week of 2024. Time-series analysis revealed consistent seasonal patterns across all regions, characterized by winter peaks (January-February) and summer troughs (July-August), as visually demonstrated in Figures 1A-C. The data revealed that the positive rate of influenza-like illness (ILI) cases exhibited significant variability during 2014-2024 (CV = 0.99 nationwide, 0.96 in southern regions, and 1.32 in northern regions). Despite these cyclical patterns, the Augmented Dickey-Fuller (ADF) test confirmed significant non-stationarity in the data, indicating inherent unpredictability in long-term trends.

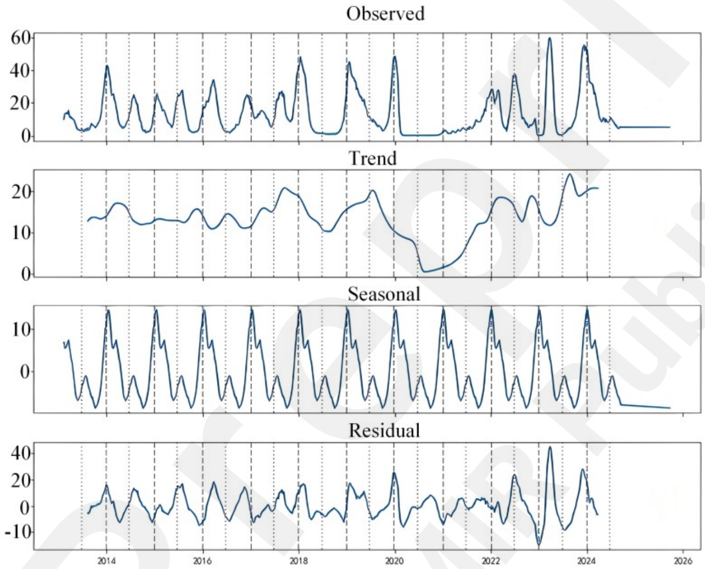
A dramatic decline in national ILI cases occurred in January 2020, coinciding with the implementation of COVID-19 containment measures including social distancing policies and mask mandates, reaching a historic trough followed by a rebound to unprecedented levels in January 2022 (Fig. 1A), which may reflect the immunity gap effect resulting from subsequent adjustments to pandemic control policies. Regional analyses showed synchronized downward trajectories: southern China exhibited a sustained decline from July 2019 to July 2020 (Fig. 1B), while northern China experienced a parallel trend starting one month earlier (June 2019-July 2020, Fig. 1C). Post-2020 decomposition revealed dual dynamics: the overall case count trend displayed steady growth with accelerated momentum from late 2020, whereas positivity rates initially declined through 2020 before resuming an upward trajectory in early 2021.

Between 2014 and 2020, seasonal extremes reliably aligned with winter peaks and summer troughs. However, the 2020-2023 period witnessed a 4-6 week delay in peak timing, a disruption attributed to pandemic-induced behavioral changes. This phase shift culminated in spring peaks during 2023 before full restoration of the original winter-summer seasonality in 2024, as comprehensively documented across Figure 1 A-C. Notably, all regions have resumed typical seasonal patterns after the third quarter of 2023, suggesting stabilization of epidemiological characteristics in the post-pandemic era.

A


Observed
Trend
Seasonal
Residual
2014
2016
2018
2020
2022
2024

B


Observed
Trend
Seasonal
Residual
2014
2016
2018
2020
2022
2024
2026

C

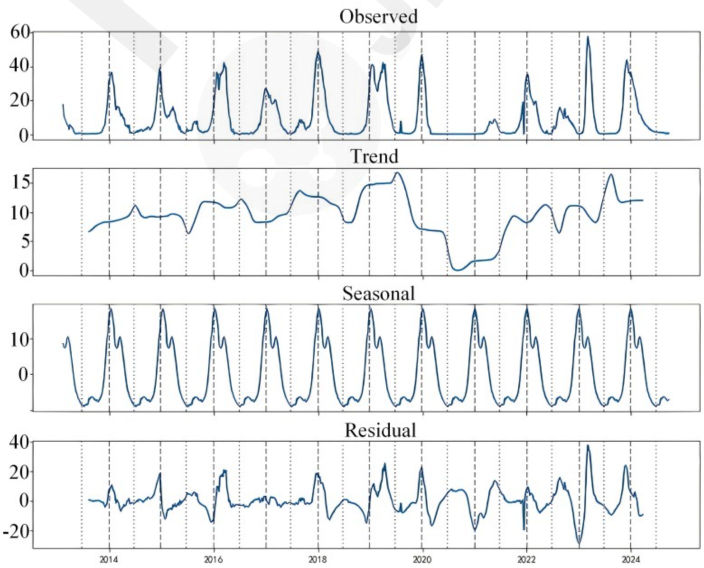

Observed
Trend
Seasonal
Residual
2014
2016
2018
2020
2022
2024

Figure 1 Seasonal decomposition of Influenza like Illness (ILI) in mainland China A: Multicycle superposition of ILI cases at national level; B: Seasonal variation patterns in southern region; C: Seasonal variation patterns in northern region.

Model Comparison

We evaluated the performance of five deep learning models-LSTM, N-BEATS, TFT, TIDE, and Transformer-in predicting the influenza - like illness (ILI) positive rate across southern, northern, and national regions of mainland China for weeks 1 through 39 of 2024. The training process of all models was monitored through fitting visualizations(Fig. 2). As shown in Figures 2A-C, the predicted values from the TIDE model closely tracked the observed ILI positive rates in the national, southern, and northern datasets, demonstrating a low mean absolute deviation on the training data. The models were assessed using four metrics: mean squared error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), with lower values indicating better predictive accuracy.

At the national level, TIDE again led in all three metrics, with an MAE of 5.551, an MSE of 43.976, and a MAPE of 72.413%, reflecting both accuracy and trend-capturing capabilities. N-BEATS had relatively larger errors, with an MAE of 9.4231, an MSE of 133.1737, and a MAPE of 105.72%. TFT had considerable deviations, marked by an MAE of 12.5381, an MSE of 235.1571, and a MAPE of 169.29%. LSTM (MAE = 6.934, MSE = 61.391, MAPE = 88.793%) demonstrated commendable performance, while Transformer (MAE = 10.6128, MSE = 132.5816, MAPE = 161.58%) had higher error rates(Fig. 3A, Table 1).

In the southern region, TIDE demonstrated the best performance, achieving an MAE of 7.554, an MSE of 89.708, and a MAPE of 74.475%. It had minimal deviation from actual values, reasonable prediction error sums, and stability across metrics. N-BEATS followed closely, with an MAE of 9.8975, an MSE of 160.9624, and a MAPE of 88.5%. In contrast, TFT underperformed with an MAE of 13.3855, an MSE of 243.3449, and a MAPE of 106.85%, showing substantial dispersion and high relative errors. LSTM (MAE = 9.460, MSE = 126.399, MAPE = 85.56%) performed well but slightly trailed TIDE, while Transformer (MAE = 11.5387, MSE = 157.0231, MAPE = 132.78%) showed intermediate results across all metrics(Fig. 3B, Table 1).

In the northern region, the performance varied. TIDE (MAE = 4.131, MSE = 28.922, MAPE = 486.087%) still outperformed other models. N - BEATS (MAE = 6.325, MSE = 58.0936, MAPE = 468.41%) showed improved performance, with stable predictions despite higher relative errors. TFT and LSTM had particularly poor results, with MAPE values reaching 2215.66% and 1918.516%, respectively. Transformer (MAE = 9.2666, MSE = 101.2513, MAPE = 1090.63%) outperformed LSTM in terms of MAE and MSE but was still affected by a high MAPE (Fig. 3C, Table 1).

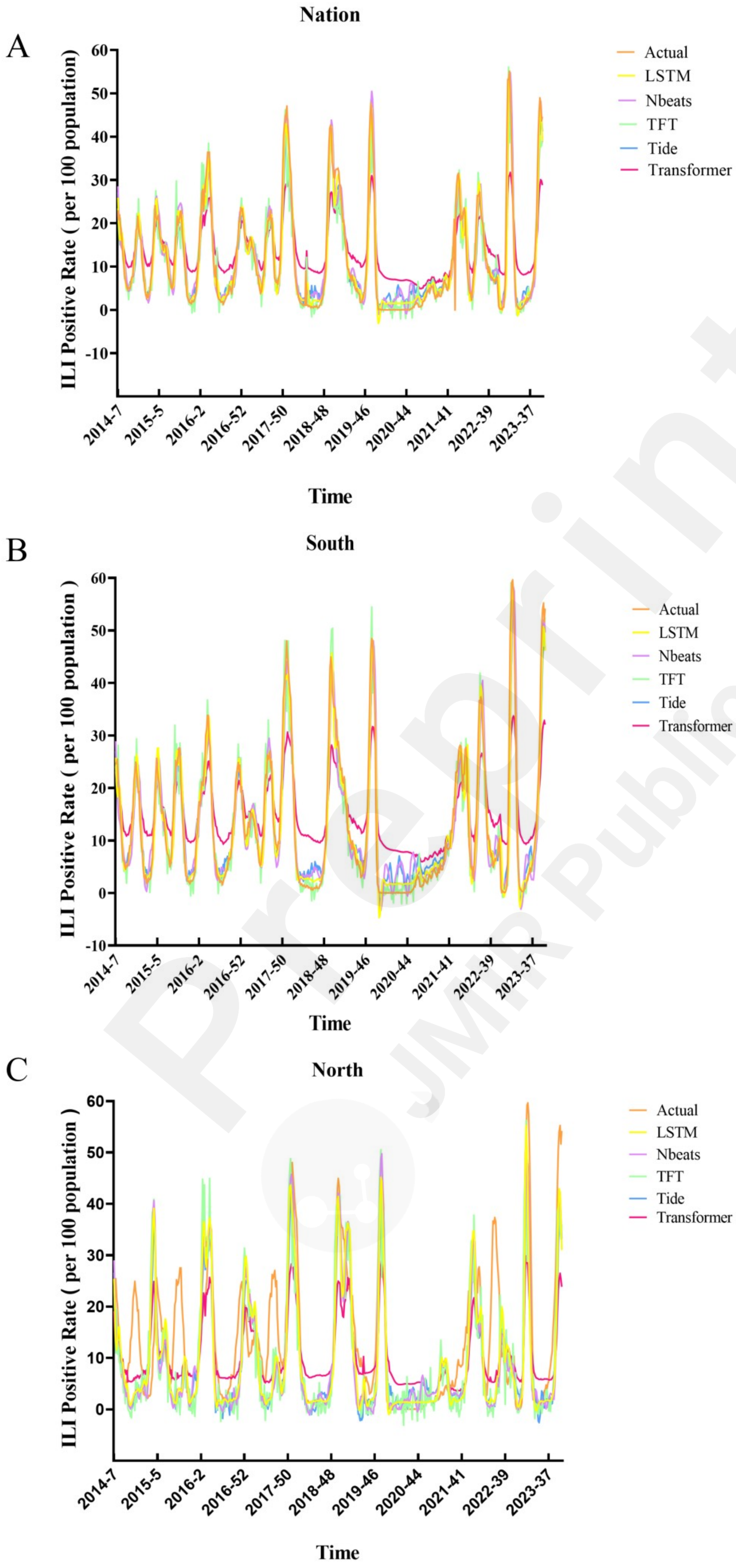
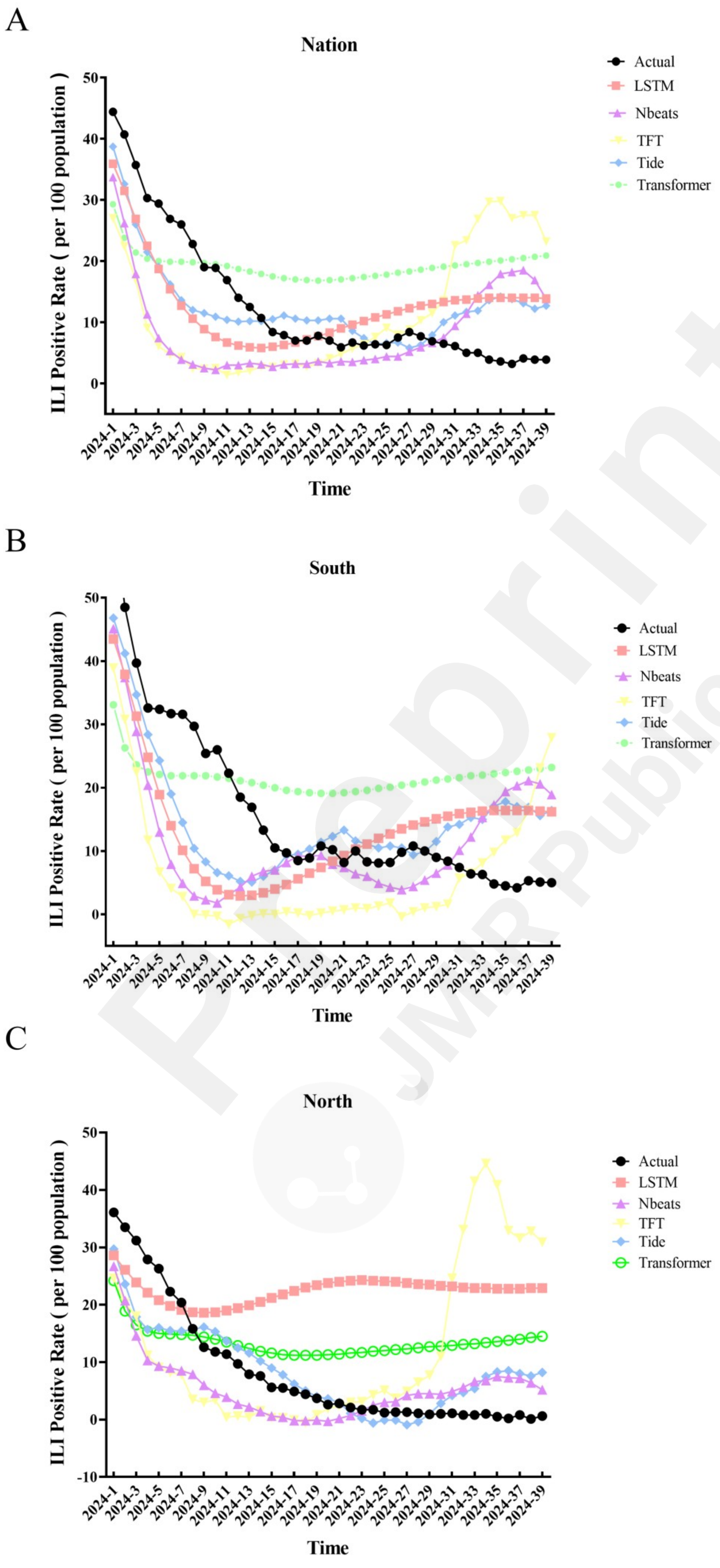

A
Nation
ILI Positive Rate (per 100 population)
Time
Actual
LSTM
Nbeats
TFT
Tide
Transformer
B
South
ILI Positive Rate (per 100 population)
Time
Actual
LSTM
Nbeats
TFT
Tide
Transformer
C
North
ILI Positive Rate (per 100 population)
Time
Actual
LSTM
Nbeats
TFT
Tide
Transformer
2014-7
2015-5
2016-2
2016-52
2017-50
2018-48
2019-46
2020-44
2021-41
2022-39
2023-37

Figure 2 The comparative forecasting trajectories of multiple predictive models alongside actual observed data across three geographical divisions: Nation (A), South (B), and North (C). (from 2014 to 2023). "Actual" denotes the real data; LSTM represents the long short-term memory neural network model; Nbeats, TFT, Tide, and Transformer stand for distinct forecasting models. Specifically, the orange line indicates the actual data, the yellow line shows the predictive results from the LSTM model, the purple line represents predictions via the Nbeats model, the green line reflects forecasts from the TFT model, the blue line is the predictive output of the Tide model, and the pink line denotes predictions by the Transformer model.



A
Nation
ILI Positive Rate (per 100 population)
Time
Actual
LSTM
Nbeats
TFT
Tide
Transformer
B
South
ILI Positive Rate (per 100 population)
Time
Actual
LSTM
Nbeats
TFT
Tide
Transformer
C
North
ILI Positive Rate (per 100 population)
Time
Actual
LSTM
Nbeats
TFT
Tide
Transformer

Figure 3 The forecasting curves of various models and the actual data series in three scenarios: Nation（A）, South（B）, and North（C）. Comparison of different forecasting models. "Actual" denotes the real data; LSTM represents the long short-term memory neural network model; Nbeats, TFT, Tide, and Transformer stand for distinct forecasting models. Specifically, the black line indicates the actual data, the pink line shows the predictive results from the LSTM model, the purple line represents predictions via the Nbeats model, the yellow line reflects forecasts from the TFT model, the blue line is the predictive output of the Tide model, and the green line denotes predictions by the Transformer model .

Table

Table1 Comparison of the fitting and prediction accuracy of the models

Region	Model	Trainning set			Test set		
		MAE	MSE	MAPE(%)	MAE	MSE	MAPE(%)
Nation	LSTM	2.022	9.341	747.490	6.934	61.391	88.793
	N-BEATS	2.468	11.296	1105.892	9.423	133.174	105.717
	TFT	2.290	12.059	564.371	12.538	235.157	169.290
	Tide	2.549	13.062	1338.139	5.551	43.976	72.413
	Transformer	5.658	47.188	3473.858	10.613	132.582	161.581
South	LSTM	2.216	10.390	573.356	9.460	126.399	85.560
	N-BEATS	3.157	18.556	1079.920	9.898	160.962	88.502
	TFT	2.134	9.599	328.874	13.385	243.345	106.853
	Tide	2.876	15.516	972.010	7.554	89.708	74.475
	Transformer	6.266	57.100	2719.575	11.539	157.023	132.782
	LSTM	2.136	9.997	1156.550	16.119	314.581	1918.516

Table 1(Continued) Comparison of the fitting and prediction accuracy of the models

Region	Model	Trainning set			Test set		
		MAE	MSE	MAPE(%)	MAE	MSE	MAPE(%)
North	N-BEATS	2.431	13.461	1740.311	6.325	58.094	468.412
	TFT	2.629	14.188	1639.950	13.610	336.891	2215.656
	Tide	2.701	15.989	2194.821	4.131	28.922	486.087
	Transformer	5.516	49.326	4291.066	9.267	101.251	1090.626

The auxiliary effects of ChatGPT

In this study, we systematically integrated ChatGPT as an auxiliary tool throughout the research workflow. This integration followed a structured approach where research tasks were formulated as standardized prompts to ensure consistency and reproducibility.

For model implementation, we developed templated prompts that specified functional requirements, architectural details, and desired outputs. When implementing complex architectures such as the Transformer model, we utilized prompts (Fig. 4). The generated code frameworks required minimal modifications before deployment, primarily related to dataset-specific preprocessing steps and hyperparameter initialization. This approach proved particularly valuable for implementing the TIDE and N-BEATS models, whose architectural complexity would otherwise necessitate extensive manual coding.

The error diagnosis process benefited substantially from ChatGPT's analytical capabilities. We established a systematic protocol wherein error messages, code context, and expected behavior were documented before submission to ChatGPT for analysis. During the implementation of a type annotation-driven data validation module, we encountered a SyntaxError: illegal target for annotation. ChatGPT pinpointed the root cause as an illegal type annotation on the right-hand side of a variable assignment statement (e.g., parsed = data["value"]: float), rather than the initially suspected class definition errors. This clear distinction between syntax rule violations and logical errors reduced debugging time from hours to seconds. Similar diagnostic success was observed when resolving memory optimization issues in the Transformer model implementation, where ChatGPT identified inefficient tensor operations that were creating computational bottlenecks (Fig. 5).

Please provide a Python script for time series forecasting using the Darts library and a Transformer model. The script should include complete steps for data loading, preprocessing, model definition, training, prediction, and evaluation. Specific requirements are as follows:

1.Use pandas to load data from a CSV file. The data should have 'Year' and 'Week' columns, which need to be converted to a datetime index.
2.Preprocess the data:
(1)Convert 'Year' and 'Week' to a datetime index using the format '%G-W%V-%u'
(2)Set the frequency to weekly, starting on Mondays
(3)Use the 'Incidence' column as the target variable
(4)Apply appropriate scaling to the data
3.Implement data splitting:
(1)Use data from 2013 to 2023 as the training set
(2)Use data from 2024 as the test set
4.Use Darts' TransformerModel for modeling:
(1)Set appropriate input and output chunk lengths (e.g., 52 weeks for yearly patterns)
(2)Configure other hyperparameters as needed
5.Perform model training and prediction
6.Calculate and print evaluation metrics such as MAE, MSE, RMSE, and MAPE
7.Use matplotlib to plot a comparison of predicted results and actual values
8.Include random seed setting to ensure reproducibility
9.Add appropriate comments explaining code functionality

The script should be well-structured, with separate functions for data loading, preprocessing, model training, prediction, and evaluation. Ensure that the code is efficient and follows best practices for time series forecasting with the Darts library.

Darts Transformer Forecasting

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from darts import TimeSeries
from darts.dataprocessing.transformers import Scaler
from darts.models import TransformerModel
from darts.metrics import mae, mse, rmse, mape
from sklearn.preprocessing import MinMaxScaler
import torch
import random

# Set random seed for reproducibility
def set_seed(seed=42):
    torch.manual_seed(seed)
```

Figure 4 Auxiliary effects of ChatGPT in model development and debugging: Example prompt template for generating Transformer model code.

"I'm encountering a SyntaxError: illegal target for annotation in my Python code when using type hints. The error occurs at line 91 where I wrote: "monitor": "train_loss", Python version: 3.11 What are the valid syntax rules for type annotations in this context, and how can I rewrite this code to comply with PEP standards while maintaining the intended type checking functionality? Please provide specific corrected code examples."

The error **"SyntaxError: illegal target for annotation"** occurs when using a type annotation incorrectly. In your case, it looks like you're trying to annotate a dictionary key, which is not a valid syntax for type hints in Python.

Understanding the Issue

Python's type hinting system (PEP 484) allows annotations for:

- Variables (`var_name: Type`)
- Function arguments (`def func(arg: Type) -> ReturnType:`)
- Class attributes (`self.attr: Type = value`)

However, **you cannot annotate dictionary keys directly like this**:

python 复制 编辑

```python
{
    "monitor": str  # ❌ SyntaxError: illegal target for annotation
}
```

Correcting the Code

Instead of directly annotating dictionary keys, use **TypedDict** from `typing` or declare the dictionary with an explicit type.

Solution 1: Using `TypedDict` (Recommended)

python 复制 编辑

```python
from typing import TypedDict

class Config(TypedDict):
    monitor: str

config: Config = {
    "monitor": "train_loss"
}
```

This approach allows precise key-level type checking and is the most PEP-compliant.

Solution 2: Using Explicit Dictionary Type Annotation

python 复制 编辑

```python
config: dict[str, str] = {
    "monitor": "train_loss"
}
```

This works but is less precise than `TypedDict` because it allows any string key, not just `"monitor"`.

Solution 3: Using a Dataclass (if appropriate)

If `"monitor"` is part of a structured configuration:

python 复制 编辑

```python
from dataclasses import dataclass

@dataclass
class Config:
    monitor: str

config = Config(monitor="train_loss")
```

This approach is useful if `monitor` belongs to a structured configuration with multiple attributes.

Figure 5 Error diagnosis workflow demonstrating ChatGPT's rapid identification and resolution of errors, significantly reducing debugging time.

Discussion

This study investigated the efficacy of five deep learning models in predicting ILI positive rates across China while exploring ChatGPT's auxiliary role in model development. Our findings demonstrate both the potential and challenges of AI-driven epidemic forecasting systems, with several important implications for public health surveillance and methodology development.

Through the plotting and analysis of the long - time series data, the significant seasonal fluctuations of ILI cases in the whole Chinese mainland and the southern and northern regions are clearly shown (Figure 1). There is a high incidence in winter (January and February), while a low trough in summer (July and August), consistent with established influenza epidemiology in temperate regions. This seasonality aligns with previous studies by Ye et al. (2019) (26) , Chen et al. (2020)(27) , Liu et al. (2023)(28), who documented similar patterns in China populations. This seasonal pattern is closely related to many factors. This seasonal pattern provides key information for subsequent model construction and prediction.

Although the time series data has an obvious seasonality, the fluctuation range is large. This phenomenon may result from the combined effects of many complex factors. The influenza virus itself is highly variable, and its antigenic drift and antigenic shift can lead to changes in the virus transmission ability, pathogenicity and population susceptibility, thus causing significant fluctuations in the number of cases (29–31).The dynamic change of the population immune level is also one of the important factors, including the natural immunity generated by previous infections, the annual differences in vaccination coverage and the phenomenon of immune escape, all of which will have an impact on the incidence of ILI(32). In addition, with the implementation of the comprehensive lifting of isolation policies in 2022, the positive rate of ILI has also decreased significantly. In terms of prevention and control measures(33), policies and behaviors such as social isolation and restriction of gatherings and mask wearing [Error: Reference source not found] have reduced population contact and virus transmission, resulting in a decrease in the number of ILI cases; in terms of medical resource allocation and testing strategies, resources have been tilted towards COVID - 19, and the testing ability and service supply of ILI have been affected, and the testing focus has shifted [Error: Reference source not found], affecting the confirmed data of ILI; changes in public behavior and health awareness, good hygiene habits reduce the risk of disease onset, and changes in medical - seeking behavior lead to statistical bias in the number of cases(34,35); the epidemic has also changed the population mobility pattern and reduced the transmission opportunity of the influenza virus. The combined action of these factors has caused fluctuations in the ILI case data, jointly leading to the instability of the data, greatly increasing the difficulty of accurately predicting the ILI trend, and posing a severe challenge to the adaptability of the prediction model.

The comparative performance of deep learning models revealed important regional differences in predictive capability. The TIDE model's superior performance nationally and in southern China suggests that its encoder-decoder architecture more effectively captures the temporal complexities of ILI transmission patterns in these regions. However, the markedly poorer performance of all models in northern China-even the relatively better-performing TIDE model-indicates fundamental challenges in modeling this region's ILI dynamics. This regional disparity likely stems from northern China's more extreme seasonal temperature variations, population density differences, and potentially greater data heterogeneity in surveillance reporting systems. These factors introduce non-linear complexities that current deep learning architectures struggle to fully capture, suggesting the need for region-specific modeling approaches.

Our integration of ChatGPT as an auxiliary research tool represents a methodological innovation in epidemiological modeling. Unlike traditional model development approaches requiring extensive programming expertise, this AI-assisted workflow not only democratizes access to sophisticated analytical techniques but also provides a reproducible framework through systematic prompt templates and error diagnosis protocols. While previous studies have explored AI assistance in clinical decision support(36–38), its application in epidemiological model development remains novel. This approach enabled parallel implementation of multiple architectures (LSTM, N-BEATS, TFT, TIDE, Transformer) within a consistent framework.

Despite these contributions, several limitations must be acknowledged. First, our data source was limited to the Chinese CDC's surveillance system, which may not capture the full spectrum of ILI cases, particularly milder cases that don't reach healthcare facilities. Second, our models primarily relied on historical ILI patterns without incorporating potentially influential covariates such as meteorological conditions, population mobility, or socioeconomic factors. Third, while ChatGPT facilitated model development, its application also introduced dependencies on proprietary systems that may affect research reproducibility. Finally, the unusual patterns during the COVID-19 pandemic created additional modeling challenges that may not persist in future influenza seasons.

Conclusion

Our study demonstrates that deep learning models, particularly TiDE, can effectively forecast regional and national ILI trends in China, while ChatGPT significantly enhances research efficiency by streamlining model development processes.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the Chinese Center for Disease Control and Prevention （CDC） repository, http://www.phsciencedata.cn.

Competing interests

The authors declare that they have no competing interests.

Funding

The study was supported by the Key Project of the Natural Science Foundation of Guangxi (2025GXNSFDA069044), Innovative Research Team Project of Guangxi Natural Science Foundation (2025GXNSFGA069002), the National Natural Science Foundation of China (NSFC; 82302550, 82460405), the Guangxi Medical University Training Program for Young Leading Talents (to Junjun Jiang). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

JJ, LY, and HL conceptualized the study, WH and WW designed the study, and BZ, XX, MZ, SL, performed the literature search. WW and WH performed resource analysis and data extraction. ZY, JL, RC provided key insights regarding data interpretation. WH, WW and JJ wrote the first draft, HL,LY edited the paper, with all authors providing critical input and edits. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

Not applicable

References

1. Spencer JA, Shutt DP, Moser SK, Clegg H, Wearing HJ, Mukundan H, et al. Distinguishing viruses responsible for influenza-like illness. J Theor Biol. 2022; 545:111145.
2. Auvinen R, Syrjänen R, Ollgren J, Nohynek H, Skogberg K. Clinical characteristics and population-based attack rates of respiratory syncytial virus versus influenza hospitalizations among adults-An observational study. Influenza Other Respir Viruses. 2022; 16(2):276–88.
3. Labella AM, Merel SE. Influenza. Med Clin North Am. 2013; 97(4):621–45,.
4. Caceres CJ, Seibert B, Cargnin Faccin F, Cardenas-Garcia S, Rajao DS, Perez DR. Influenza antivirals and animal models. FEBS Open Bio. 2022; 12(6):1142–65.
5. Lu FS, Hou S, Baltrusaitis K, Shah M, Leskovec J, Sosic R, et al. Accurate Influenza Monitoring and Forecasting Using Novel Internet Data Streams: A Case Study in the Boston Metropolis. JMIR Public Health Surveill. 2018; 4(1):e4.
6. Kim J, Ahn I. Weekly ILI patient ratio change prediction using news articles with support vector machine. BMC Bioinformatics. 2019; 20(1):259.
7. Qualls N, Levitt A, Kanade N, Wright-Jegede N, Dopson S, Biggerstaff M, et al. Community Mitigation Guidelines to Prevent Pandemic Influenza - United States, 2017. MMWR Recomm Rep. 2017;

66(1):1–34.
8. Nsoesie EO, Oladeji O, Abah ASA, Ndeffo-Mbah ML. Forecasting influenza-like illness trends in Cameroon using Google Search Data. Sci Rep. 2021; 11(1):6713.
9. Ohkusa Y, Shigematsu M, Taniguchi K, Okabe N. Experimental surveillance using data on sales of over-the-counter medications--Japan,November 2003-April 2004. MMWR Suppl. 2005; 54:47–52.
10. Vergu E, Grais RF, Sarter H, Fagot JP, Lambert B, Valleron AJ, et al. Medication sales and syndromic surveillance, France. Emerg Infect Dis. 2006; 12(3):416–21.
11. Greff K, Srivastava RK, Koutnik J, Steunebrink BR, Schmidhuber J. LSTM: A Search Space Odyssey. IEEE Trans Neural Netw Learn Syst. 2017; 28(10):2222–32.
12. Oreshkin B N, Carpov D, Chapados N, Bengio Y. N-BEATS: neural basis expansion analysis for interpretable time series forecasting. ARXIV. 2019; 1:1–31.
13. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. , N. Attention Is All You Need. arXiv preprint. 2023;
14. Farooq A, Irfan Uddin M, Adnan M, Alarood AA, Alsolami E, Habibullah S. Interpretable multi-horizon time series forecasting of cryptocurrencies by leverage temporal fusion transformer. Heliyon. 2024 ;10(22):e40142.
15. Das, A., Kong, W., Leach, A., Mathur, S., Sen, R., Yu, R. Long-term Forecasting with TiDE: Time-series Dense Encoder. arXiv preprint. 2024;
16. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011; 343:d5928.
17. Cheng Y, Bai Y, Yang J, Tan X, Xu T, Cheng R. Analysis and prediction of infectious diseases based on spatial visualization and machine learning. Sci Rep. 2024; 14(1):28659.
18. Faes L, Wagner SK, Fu DJ, Liu X, Korot E, Ledsam JR, et al. Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. Lancet Digit Health. 2019; 1(5):e232–42.
19. Chen J, See KC. Artificial Intelligence for COVID-19: Rapid Review. J Med Internet Res. 2020; 22(10):e21476.
20. Kim JK, Chua M, Rickard M, Lorenzo A. ChatGPT and large language model (LLM) chatbots: The current state of acceptability and a proposal for guidelines on utilization in academic medicine. J Pediatr Urol. 2023; 19(5):598–604.
21. Levin G, Brezinov Y, Meyer R. Exploring the use of ChatGPT in OBGYN: a bibliometric analysis of the first ChatGPT-related publications. Arch Gynecol Obstet. 2023; 308(6):1785–9.
22. Ignjatović A, Stevanović L. Efficacy and limitations of ChatGPT as a biostatistical problem-solving tool in medical education in Serbia: a descriptive study. J Educ Eval Health Prof. 2023; 20:28.
23. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med. 2019; 25(1):24–9.
24. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med. 2023; 29(8):1930–40.
25. China National Influenza Center Database. https://ivdc.chinacdc.cn/cnic/. Accessed 30 September 2024.
25. Ye C, Zhu W, Yu J, Li Z, Zhang Y, Wang Y, et al. Understanding the complex seasonality of seasonal influenza A and B virus transmission: Evidence from six years of surveillance data in Shanghai, China. Int J Infect Dis. 2019; 81:57–65.
26. Chen H, Xiao M. Seasonality of influenza-like illness and short-term forecasting model in Chongqing from 2010 to 2022. BMC Infect Dis. 2024; 24(1):432.
27. Liu X, Peng Y, Chen Z, Jiang F, Ni F, Tang Z, et al. Impact of non-pharmaceutical interventions during COVID-19 on future influenza trends in Mainland China. BMC Infect Dis. 2023; 23(1):632.
28. Nelson MI, Holmes EC. The evolution of epidemic influenza. Nat Rev Genet. 2007; 8(3):196–205.
29. Moya A, Holmes EC, González-Candelas F. The population genetics and evolutionary epidemiology of RNA viruses. Nat Rev Microbiol. 2004; 2(4):279–88.
30. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. The genomic and epidemiological dynamics of human influenza A virus. Nature. 2008; 453(7195):615–9.
31. Needle RF, Russell RS. Immunity Debt, a Gap in Learning, or Immune Dysfunction? Viral Immunol. 2023; 36(1):1–2.

32. Davis WW, Mott JA, Olsen SJ. The role of non-pharmaceutical interventions on influenza circulation during the COVID-19 pandemic in nine tropical Asian countries. Influenza Other Respir Viruses. 2022; 16(3):568–76.
33. Feng L, Zhang T, Wang Q, Xie Y, Peng Z, Zheng J, et al. Impact of COVID-19 outbreaks and interventions on influenza in China and the United States. Nat Commun. 2021; 12(1):3249.1
34. Hazra D, Chandy GM, Thanjavurkar A, Gunasekaran K, Nekkanti AC, Pal R, et al. A clinico-epidemiological profile, coinfections and outcome of patients with Influenza Like Illnesses (ILI) presenting to the emergency department during the COVID-19 pandemic. J Family Med Prim Care. 2023; 12(4):672–8.
35. Arslan S. Exploring the Potential of Chat GPT in Personalized Obesity Treatment. Ann Biomed Eng. 2023; 51(9):1887–8.
36. Giske CG, Bressan M, Fiechter F, Hinic V, Mancini S, Nolte O, et al. GPT-4-based AI agents-the new expert system for detection of antimicrobial resistance mechanisms? J Clin Microbiol. 2024; 62(11):e0068924.
37. Siyao L, Yu W, Gu T, Lin C, Wang Q, Qian C, et al. Bailando++: 3D Dance GPT With Choreographic Memory. IEEE Trans Pattern Anal Mach Intell. 2023; 45(12):14192–207.

Supplementary Files

Untitled.
URL: http://asset.jmir.pub/assets/b7b5a7ed29e0cf6dd898a2232657f140.pdf

Untitled.
URL: http://asset.jmir.pub/assets/5590ddb8bf02ebff8757d0bde83503a0.xlsx

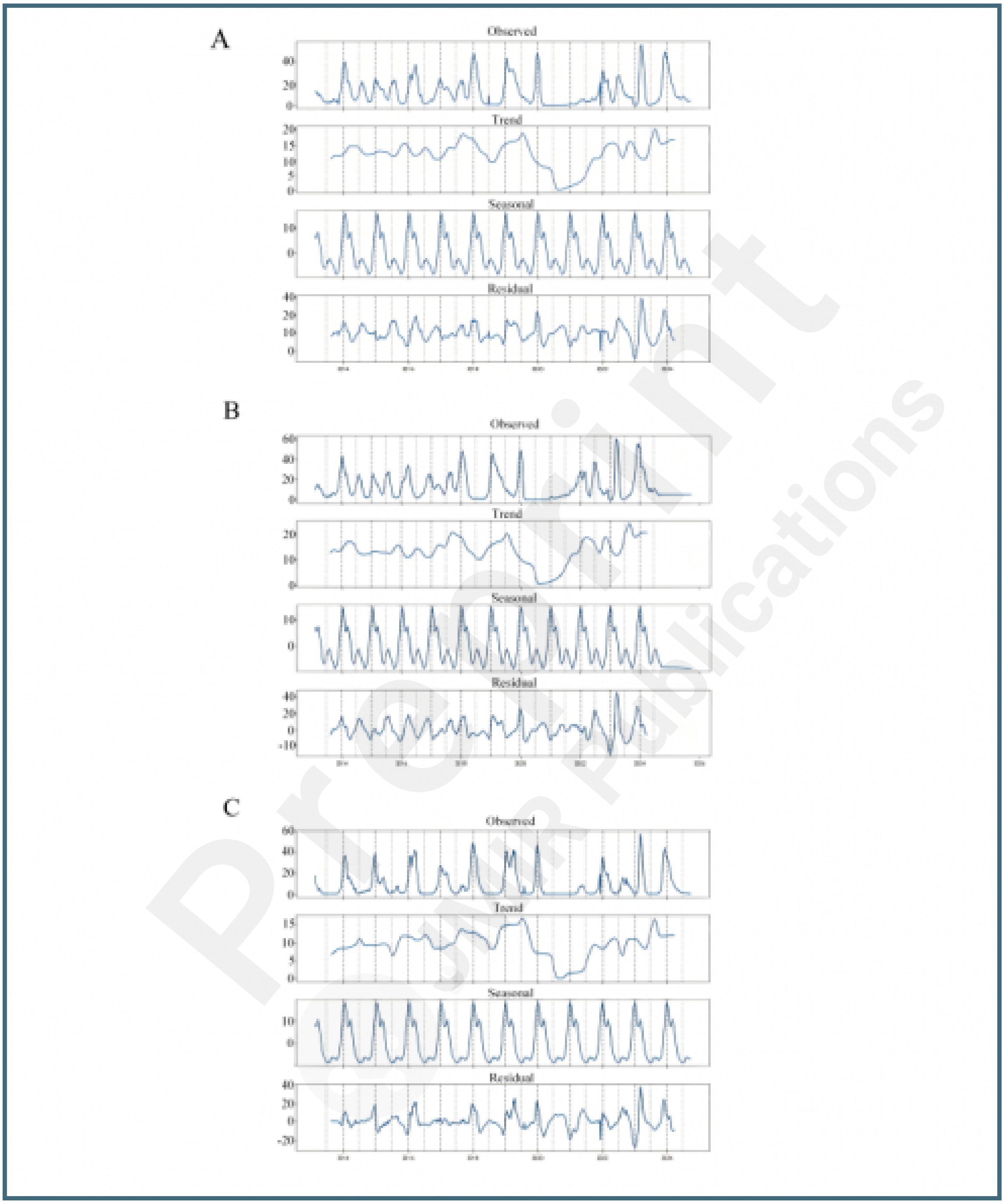
Untitled.
URL: http://asset.jmir.pub/assets/c80608a27525f570e4b077bee5cc6ccf.xlsx

?????s

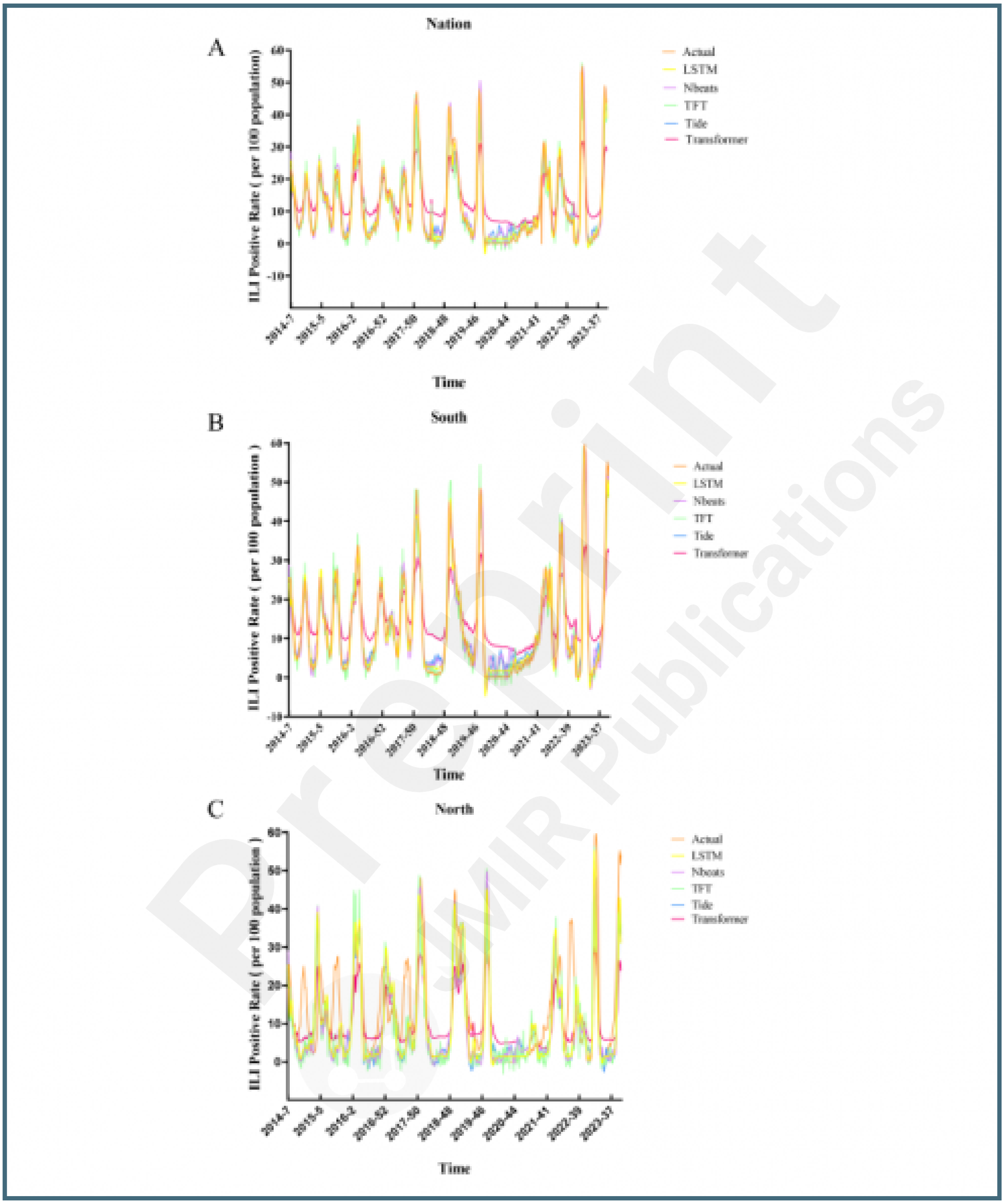
Untitled.
URL: http://asset.jmir.pub/assets/18bfd443212b9a1a717b0d9ff41fc667.docx

??? TOC/Feature ??s

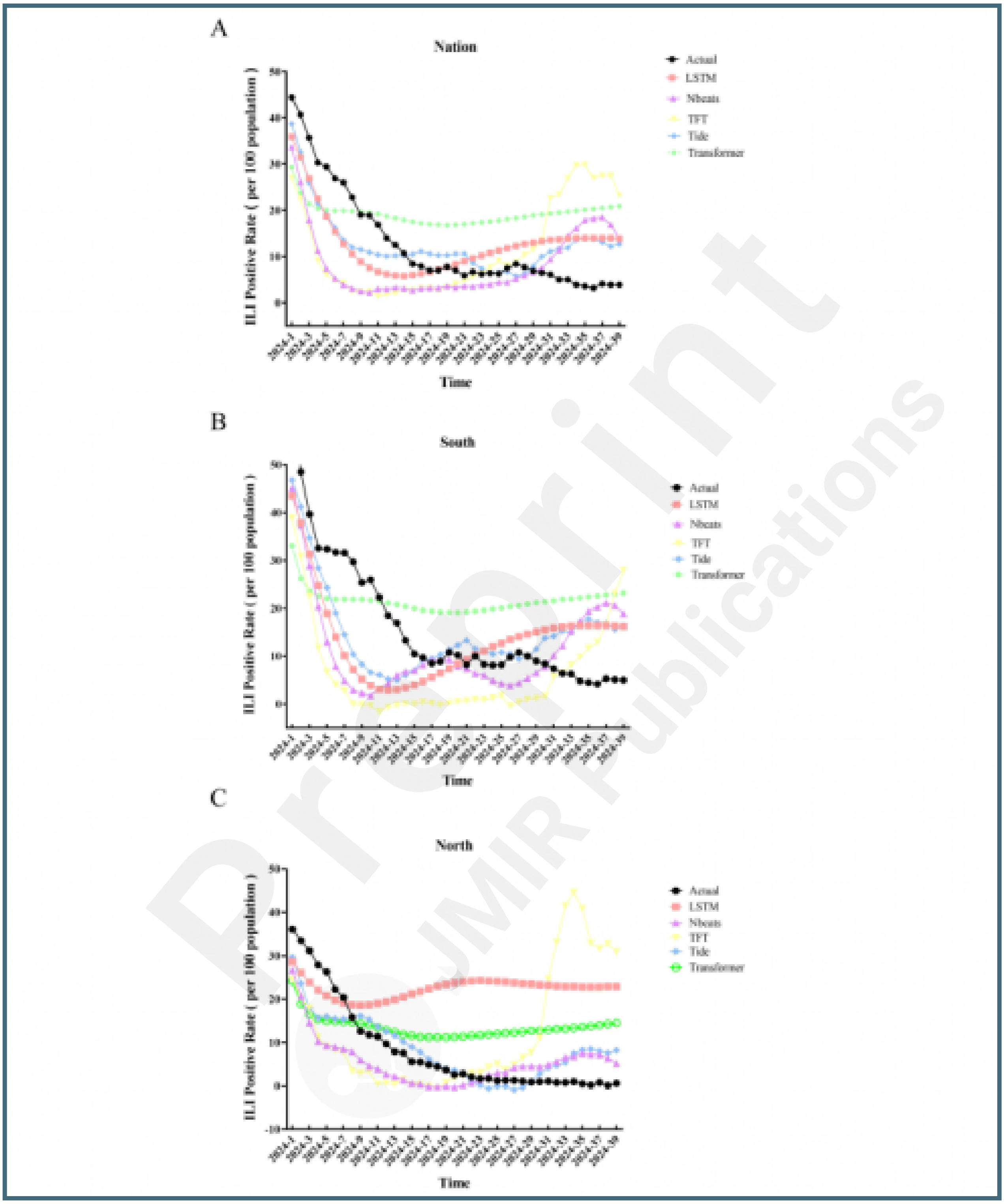
Untitled.


A
Observed
Trend
Seasonal
Residual
B
Observed
Trend
Seasonal
Residual
C
Observed
Trend
Seasonal
Residual

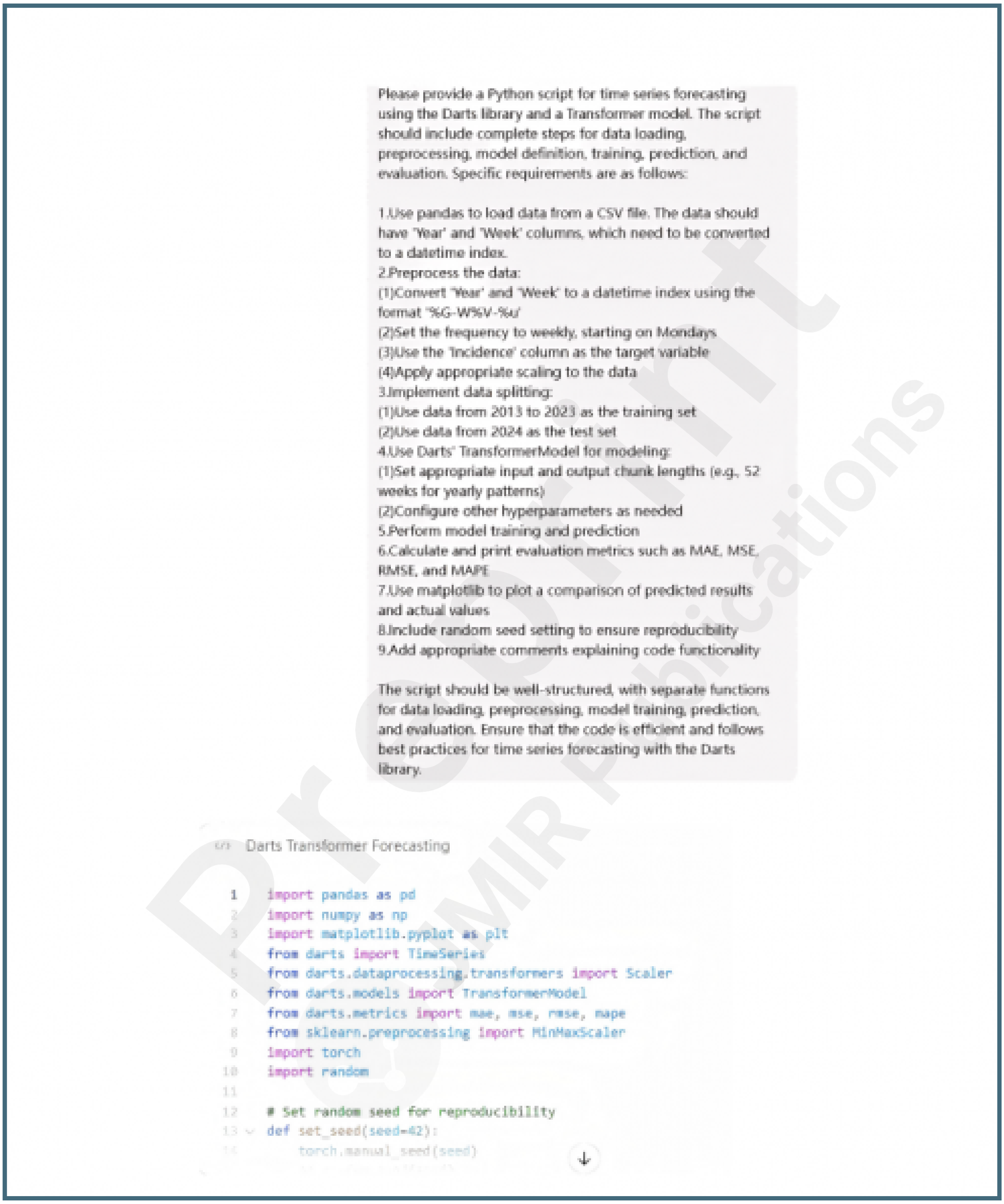
Untitled.


Nation
A
ILI Positive Rate (per 100 population)
Time
Actual
LSTM
Nbeats
TFT
Tide
Transformer
South
B
ILI Positive Rate (per 100 population)
Time
Actual
LSTM
Nbeats
TFT
Tide
Transformer
North
C
ILI Positive Rate (per 100 population)
Time
Actual
LSTM
Nbeats
TFT
Tide
Transformer

Untitled.


A
Nation
Actual
LSTM
Nbeats
TFT
Tide
Transformer
ILI Positive Rate (per 100 population)
Time
B
South
Actual
LSTM
Nbeats
TFT
Tide
Transformer
ILI Positive Rate (per 100 population)
Time
C
North
Actual
LSTM
Nbeats
TFT
Tide
Transformer
ILI Positive Rate (per 100 population)
Time

Untitled.



Please provide a Python script for time series forecasting using the Darts library and a Transformer model. The script should include complete steps for data loading, preprocessing, model definition, training, prediction, and evaluation. Specific requirements are as follows:

1.Use pandas to load data from a CSV file. The data should have 'Year' and 'Week' columns, which need to be converted to a datetime index.
2.Preprocess the data:
(1)Convert 'Year' and 'Week' to a datetime index using the format '%G-W%V-%u'
(2)Set the frequency to weekly, starting on Mondays
(3)Use the 'Incidence' column as the target variable
(4)Apply appropriate scaling to the data
3.Implement data splitting:
(1)Use data from 2013 to 2023 as the training set
(2)Use data from 2024 as the test set
4.Use Darts' TransformerModel for modeling:
(1)Set appropriate input and output chunk lengths (e.g., 52 weeks for yearly patterns)
(2)Configure other hyperparameters as needed
5.Perform model training and prediction
6.Calculate and print evaluation metrics such as MAE, MSE, RMSE, and MAPE
7.Use matplotlib to plot a comparison of predicted results and actual values
8.Include random seed setting to ensure reproducibility
9.Add appropriate comments explaining code functionality

The script should be well-structured, with separate functions for data loading, preprocessing, model training, prediction, and evaluation. Ensure that the code is efficient and follows best practices for time series forecasting with the Darts library.

Darts Transformer Forecasting

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from darts import TimeSeries
from darts.dataprocessing.transformers import Scaler
from darts.models import TransformerModel
from darts.metrics import mae, mse, rmse, mape
from sklearn.preprocessing import MinMaxScaler
import torch
import random

# Set random seed for reproducibility
def set_seed(seed=42):
    torch.manual_seed(seed)
```

Untitled.

"I'm encountering a SyntaxError: illegal target for annotation in my Python code when using type hints. The error occurs at line 91 where I wrote: "monitor": "train_loss", Python version: 3.11 What are the valid syntax rules for type annotations in this context, and how can I rewrite this code to comply with PEP standards while maintaining the intended type checking functionality? Please provide specific corrected code examples."

The error "**SyntaxError: illegal target for annotation**" occurs when using a type annotation incorrectly. In your case, it looks like you're trying to annotate a dictionary key, which is not a valid syntax for type hints in Python.

Understanding the Issue

Python's type hinting system (PEP 484) allows annotations for:

- Variables (`var_name: type`)
- Function arguments (`def func(arg: type) -> ReturnType:`)
- Class attributes (`self.attr: type = value`)

However, **you cannot annotate dictionary keys directly like this**:

```python
{
    "monitor": str  # ❌ SyntaxError: illegal target for annotation
}
```

Correcting the Code

Instead of directly annotating dictionary keys, use **TypedDict** from `typing` or declare the dictionary with an explicit type.

Solution 1: Using `TypedDict` (Recommended)

```python
from typing import TypedDict

class Config(TypedDict):
    monitor: str

config: Config = {
    "monitor": "train_loss"
}
```

This approach allows precise key-level type checking and is the most PEP-compliant.

Solution 2: Using Explicit Dictionary Type Annotation

```python
config: dict[str, str] = {
    "monitor": "train_loss"
}
```

This works but is less precise than `TypedDict` because it allows any string key, not just `"monitor"`.

Solution 3: Using a Dataclass (if appropriate)

If `"monitor"` is part of a structured configuration:

```python
from dataclasses import dataclass

@dataclass
class Config:
    monitor: str

config = Config(monitor="train_loss")
```

This approach is useful if `monitor` belongs to a structured configuration with multiple attributes.