Reviewed Preprint
v1 • September 11, 2025
Not revised

Reviewed Preprint
v2 • September 3, 2026
Revised by authors

For correspondence:
hanxiao@fzu.edu.cn
xiongsheng@jnu.edu.cn
ptcdc001@163.com
yuanyecat@vip.sina.com

# These authors contributed equally to this work

Competing interests: No competing interests declared

Funding: See page 31

Reviewing editor: George Okoli, University of Hong Kong, Hong Kong

Meteorological Drivers of Influenza A and B Positivity in a Subtropical Chinese City: A Six-Year Surveillance Study Integrating Distributed Lag Non-Linear Models and Deep Learning

Long Xie[1,#], Meng-Jie Zhang[2,#], Jin-Lin Tan[3,4,#], Yi-Xin Ling[2], Zhe-Qiang Xue[3,4], Zhi-Shen Wu[3,4], Jun-Ju Huang[3,4], Jian-Ling Chen[2], Ze-Fan Ruan[3,4], Jing Qian[2], Hai-Yong Pan[2], Xiao Han[5,6] ✉, Sheng Xiong[7] ✉, Long-Mei Ling[2] ✉, Xi-Wen Jiang[3,4,8] ✉

[1]Translational & Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom • [2]Department of Microbiology Laboratory, Putian Centre for Disease Control and Prevention, Putian, China • [3]Research Institute, DAAN Gene Co., Ltd, Guangzhou, China • [4]The Medicine and Biological Engineering Technology Research Centre of the Ministry of Health, Guangzhou, China • [5]College of Biological Science and Engineering, Fuzhou University, Fuzhou, China • [6]Department of Infection, Mengchao Hepatobiliary Hospital of Fujian Medical University, Fuzhou, China • [7]College of Life Science and Technology, Jinan University, Guangzhou, China • [8]School of Life Sciences and Biopharmaceutics, Guangdong Pharmaceutical University, Guangzhou, China

eLife Assessment

This is a **valuable** contribution to influenza surveillance. After revisions, addressing methodological and reporting issues, the claims of providing an early warning tool now align more with the reported study results, and ultimately, the evidence presented provide a **solid** addition to the evidence base.

https://doi.org/10.7554/eLife.107767.2.sa2

Abstract

Influenza transmission in subtropical regions is shaped by complex, non-linear meteorological conditions, while surveillance-based forecasting has been further complicated by fluctuating testing intensity and non-pharmaceutical interventions during the COVID-19 period. Existing approaches often either characterize lagged environmental associations without forecasting capacity or apply deep learning without clear epidemiological interpretability. Using six years (2018–2023) of multi-site influenza surveillance data from Putian, a subtropical coastal city in southeastern China, we developed a two-stage framework to characterize meteorological associations and improve short-horizon forecasting. Influenza positivity rates were used as the primary outcome to mitigate testing-related surveillance bias. Distributed lag non-linear models (DLNMs) were constructed to estimate subtype-specific, lagged associations between meteorological variables and influenza positivity, and a Bayesian-optimized long short-term memory (LSTM) network integrating meteorological, autoregressive, and socio-behavioral covariates was developed to forecast influenza A and B activity. Influenza A positivity was associated with warmer conditions, peaking at 15°C (relative risk [RR] = 3.01, 95% CI: 1.03-8.78) and narrow diurnal temperature ranges, whereas influenza B positivity was associated with colder (8°C; RR = 26.12, 95% CI: 7.79-87.61), more humid (91%; RR = 1.52, 95% CI: 1.00-2.29), and lower-radiation conditions. The LSTM showed lower forecast error than covariate-matched autoregressive baselines for both subtypes (Mean Absolute Error [MAE]: 0.009 vs. 0.136 for

influenza A, and 0.002 vs. 0.049 for influenza B; Symmetric Mean Absolute Percentage Error [SMAPE]: 0.521 vs. 1.212, and 0.484 vs. 0.810, respectively). Covariate-ablation and SHapley Additive exPlanations (SHAP) analyses suggested that historical positivity and testing-related variables contributed materially to forecast stability under surveillance variability. External evaluation in Sanming provided preliminary evidence of model portability. This integrated DLNM-LSTM framework offers an interpretable approach for characterizing subtype-specific meteorological associations and improving influenza forecasting in subtropical settings.

Importance

Influenza imposes substantial global morbidity and mortality each year, and forecasting is particularly difficult in subtropical regions, where transmission occurs year-round, meteorological drivers are non-linear and lagged, and the COVID-19 period introduced pronounced epidemiological non-stationarity. Forecasts based on raw case counts are further affected by surveillance-related bias, as counts partly reflect testing intensity and clinician vigilance rather than underlying transmission alone. Existing work has largely applied inferential and deep-learning approaches in isolation, limiting either interpretability or forecasting flexibility. We address these issues with a two-stage DLNM–LSTM framework that uses influenza positivity rates as the primary endpoint and incorporates weekly detection volumes, mask-wearing stringency indices, day-and-week and non-local population proportion as socio-behavioral covariates, in order to mitigate testing-related surveillance bias and accommodate pandemic-era non-stationarity. The framework reveals distinct, subtype-specific meteorological associations for influenza A and B, shows lower forecast error than a covariate-matched autoregressive baseline, and provides preliminary evidence of portability to a second subtropical city. It offers an interpretable, climate-informed methodological foundation for future operational surveillance developments in subtropical influenza forecasting.

Introduction

Influenza imposes a substantial global health burden, with an estimated 290,000–650,000 respiratory deaths and 3–5 million severe cases among one billion infections annually (Iuliano et al. 2018). In China alone, influenza is associated with an estimated 88,100 excess respiratory deaths per year (Li et al. 2019), with children, the elderly, and individuals with chronic comorbidities disproportionately affected. Effective preparedness therefore requires both a mechanistic understanding of the drivers of influenza activity and the operational capacity to forecast its short-term trajectory. Because influenza A and B viruses are transmitted primarily through respiratory droplets and contact (Uyeki et al. 2022), with transmission dynamics principally influenced by viral characteristics, host susceptibility, and environmental conditions (Javanian et al. 2021), meteorological variables, particularly temperature, humidity, and precipitation, represent a critical, potentially modifiable dimension of transmission (Moriyama, Hugentobler, and Iwasaki 2020).

Meteorological drivers of influenza transmission have been extensively characterized in temperate settings, where a pronounced winter peak (Ryu and Cowling 2021) is associated with low ambient temperatures (5-8°C) that stabilize aerosolized influenza virions and impair host mucociliary clearance (Moriyama, Hugentobler, and Iwasaki 2020), and with low absolute humidity (<40%) that prolongs airborne aerosol residence while compromising respiratory epithelial defenses (Peci et al. 2019; Lowen and Steel 2014; Eccles 2002; Tamerius et al. 2013). However, these temperate-zone patterns do not generalize to subtropical regions, where influenza circulates year-round, exhibits multiple annual peaks, and demonstrates meteorological coupling that differs substantially in directionality and lag structures from temperate observations (Deyle et al. 2016; Soebiyanto et al. 2015). Despite the demographic and climatic significance of subtropical regions for global influenza dynamics (Bloom-Feshbach et al. 2013; Tamerius et al. 2013), dedicated quantitative studies in subtropical Chinese urban areas, particularly in non-megacity settings that nonetheless harbor substantial populations and represent prevailing

regional urbanization patterns, remain sparse. This geographic and methodological gap leaves subtropical influenza preparedness without the region-specific evidence required for precision-oriented public health planning.

A key reason subtropical weather-influenza relationships remain difficult to characterize is their non-linear, threshold-dependent, and lagged nature. Conventional linear analyses, such as logistic regression, assume simple contemporaneous relationships and thus fail to capture these dynamics with the precision required for epidemiological inference (Ng et al. 2022). Moreover, meteorological effects on influenza transmission also unfold over days to weeks, further complicating association analysis (Ng et al. 2022; Si et al. 2024). Distributed Lag Non-linear Models (DLNM) provide a framework for jointly analyzing exposure-response and lag-response relationships (Gasparrini, Armstrong, and Kenward 2010; Gasparrini 2011). DLNM has been applied in empirical analyses of influenza transmission patterns and in lagged risk analyses between influenza and meteorological factors (Liu et al. 2020; Guo et al. 2019). However, DLNM is fundamentally an inferential tool designed for retrospective association characterization and does not generate operational forecasts capable of supporting prospective public health decision-making.

On the forecasting side, traditional statistical models, such as Autoregressive Integrated Moving Average (ARIMA) approaches, despite their widespread use for influenza prediction (Wang et al. 2017), struggle to capture non-linear, lagged interactions among multiple meteorological drivers (Shoji, Katayama, and Sano 2011). Existing studies often overlook spatiotemporal heterogeneity intrinsic to subtropical contexts (Zhang et al. 2022). Although time series analysis approaches and deep neural learning architectures, such as Recurrent Neural Networks (RNNs) (Venna et al. 2018), can accommodate non-linear temporal patterns (Venna et al. 2018), their performance is hampered when baseline algorithms are not deeply customized, due to gradient vanishing or explosion and memory decay over longer horizons (Liu et al. 2021). Long Short-Term Memory (LSTM) algorithms, an advanced RNN variant, address these issues through gating mechanisms that enable learning of long-term temporal dependencies (Absar et al. 2022; Liu et al. 2021), and have shown potential in influenza forecasting (Absar et al. 2022), although prior applications have often struggled to architectural complexity, high computational cost, and typically lower interpretability than regression-based methods (Liu et al. 2021). For instance, LSTM algorithms have demonstrated good predictive performance in the Chinese subtropical Fujian region but remained limited to identify subtype-specific virological differences (Zhu et al. 2022). Another study proposed a complex seasonal autoregressive integrated moving average-LSTM (SARIMA-LSTM) hybrid utilizing singular spectrum analysis to predict influenza in Shanxi. However, its analytical complexity and computational intensity limited the explicit incorporation of confounding behavioral variables (Zhao et al. 2023).

Reviewing these literatures, three substantive gaps persist at the intersection of meteorological characterization and forecasting for subtropical influenza. First, most existing studies apply DLNM or LSTM in isolation, foregoing the potential synergy between interpretable association modeling and flexible non-linear forecasting. Second, forecasting models trained on data spanning the Corona virus disease 2019 (COVID-19) period rarely account explicitly for the epidemiological non-stationarity introduced by non-pharmaceutical interventions (NPIs) and fluctuating testing intensity, which can distort case-based signals in subtype-specific ways. Third, the choice of modeling endpoint, typically raw case counts rather than positivity, rendering them highly vulnerable to surveillance intensity biases, as counts confound underlying transmission with testing behavior, particularly during the pandemic and post-pandemic period.

To address these gaps, we developed a dual-stage analytical framework leveraging six years (2018-2023) of multi-site influenza surveillance data and detailed meteorological records from Putian, a representative subtropical coastal city in southeastern China that exemplifies the under-investigated non-megacity urban context described above. The 2018–2023 window deliberately spans a structural epidemiological transition: two pre-pandemic baseline years (2018–2019), three NPI-modulated years (2020–2022), and one post-suppression rebound year (2023). In the first stage, DLNMs characterize subtype-specific, non-linear, and lag-distributed associations between

meteorological variables and influenza A and B positivity. In the second stage, a Bayesian-optimized LSTM network integrating meteorological, autoregressive, and socio-behavioral covariates is used for short-horizon forecasting, benchmarked against a covariate-matched multivariate ARIMA model and evaluated in an independent subtropical city (Sanming) as a preliminary test of model portability. Influenza positivity rate is used as the primary modeling endpoint to mitigate testing-related surveillance bias. Our aim is to provide an interpretable, climate-informed forecasting approach for subtropical influenza that can serve as a methodological foundation for future operational surveillance developments.

Methods

Ethics statement

This study was institutionally approved by the Research Ethics Committee of the Putian Centre for Disease Control and Prevention (Putian CDC, approval number: Ethics Review 2020-003), Putian, and the Medicine and Biological Engineering Technology Research Centre of the Ministry of Health (MBETRC), Guangzhou, China. All procedures complied with the later amendments of the 1964 Declaration of Helsinki and relevant ethical guidelines for biomedical research. A major methodological strength of this study lies in its robust, uninterrupted longitudinal data collection framework spanning January 1, 2018, to December 31, 2023. While the analytical timeline encompasses a “retrospective” phase (January 1, 2018 – October 13, 2020) prior to formal ethical approval, and a “prospective” phase thereafter, we emphasize that this distinction represents a purely administrative demarcation regarding the timing of ethical approval. It does not reflect any shift in demographic cohorts, sentinel hospital locations, or data collection methodologies. Importantly, the historical data (2018–2020) were not subjected to the recall biases or misclassification risks typical of traditional retrospective chart reviews. Rather, they were systematically extracted from a continuously operating, highly standardized public health sentinel surveillance network. From the inception of data collection through the end of 2023, the local CDC maintained absolute uniformity in clinical influenza-like illness (ILI) definitions, nasopharyngeal swabbing procedures, and real-time reverse transcription polymerase chain reaction (RT-PCR) diagnostic assays. Consequently, the pre-2020 data possess the high-fidelity characteristics of a strict prospective cohort, ensuring unparalleled longitudinal consistency and mitigating temporal measurement bias across the entire pre-pandemic, pandemic, and post-restriction timeline.

Rigorous procedures were employed to ensure patient confidentiality and data privacy. Influenza surveillance data were fully anonymized prior to being accessed by the research team. The dataset included only essential epidemiological information (such as onset date, gender, age, and residential district) necessary for the analysis, excluding any personal identifiable information. Additionally, individual-level data on vaccination status or detailed socioeconomic profiles were not routinely collected. Any potential linkage between coded data and original patient identifiers maintained by Putian CDC was strictly inaccessible to the researchers. Consequently, the Research Ethics Committee waived the requirement for individual informed consent. Throughout the data collection process, the diagnosis of all ILI cases, information and specimen collection, and influenza virus nucleic acid testing strictly followed prevailing national standards, including the Diagnostic Criteria for Influenza (source: National Health Commission of the People’s Republic of China) and the National Influenza Surveillance Technical Specifications (source: China CDC). All participating medical professionals, CDC personnel, and laboratory technicians underwent professional training to ensure research quality. Meteorological and COVID-19 pandemic-related covariate data were sourced from publicly available databases, ensuring no privacy concerns. The anonymized data were securely managed, used solely for the stated research purposes, and were not disclosed to any third party.

Study area

Putian city, on China's southeast coast (latitudes 24°59'N–25°46'N; longitudes 118°27'E–119°39'E), is an exemplary site for examining the impact of meteorological factors on influenza transmission in subtropical regions. It experiences a subtropical monsoon climate with warm, humid summers and mild, dry winters. The annual mean temperature ranges 18–21°C, with summer highs up to 38°C and winter lows at 0.4°C. Annual precipitation averages 1538.8 mm, with 72% occurring from April to September. Covering 4200 km² and comprising five counties/districts with a population exceeding 3.2 million, Putian is a commercial/manufacturing hub with high population density and robust transport infrastructure, promoting mobility and social interactions that drive influenza transmission. Since 2004, Putian CDC has operated a comprehensive influenza surveillance network, meeting national standards, employing standardized RT-PCR testing for influenza A/B since late 2017, supported by a trained workforce, ensuring high-quality longitudinal data for robust analyses. Together, Putian's climatic variability, socio-economic characteristics, and surveillance capacity position it as a representative site for this study, with insights potentially generalizable to other subtropical settings with similar environmental and epidemiological traits. Sanming city, between latitudes 25°30'N and 27°97'N and longitudes 116°22'E and 118°39'E, is a mountainous subtropical area about 330 km from Putian and was selected as an external validation data source.

Study design

This collaborative study by Putian CDC and MBETRC utilized a prospective cohort design, with retrospective analysis of historical data to systematically assess non-linear quantitative relationships between multiple meteorological factors and both influenza A and B infection risks, ultimately constructing a predictive network for outbreaks. The technical workflow is illustrated in Figure 1.

The study spanned January 1, 2018, to December 31, 2023, integrating four categories of data sources: (i) ILI and laboratory-confirmed cases from seven influenza sentinel hospitals across Putian's urban and rural areas, ensuring representative coverage of diverse healthcare-seeking populations; (ii) daily meteorological data; (iii) COVID-19 associated public health intervention indicators (mask-wearing stringency indices) recorded from January 2020 onwards; and (iv) demographic mobility indicators (non-local population proportion) obtained from ILI consultation records. To ensure consistency between meteorological measurements and influenza incidence records across all data sources, we applied rigorous quality control and pre-processing procedures, including: temporal alignment of all data streams to a unified daily resolution; missing value imputation using temporally adjacent observations for sporadic gaps (<5% of records); cross-validation of laboratory-confirmed cases against ILI consultation records to identify and resolve coding inconsistencies; and standardization of meteorological measurements against the regional monitoring network's established calibration protocols. Final datasets underwent independent verification by two co-investigators to ensure analytical reliability.

DLNM was first constructed to screen meteorological factors and other covariates with substantial influence on influenza seasonality. Subsequently, an LSTM neural network was developed within the same covariate system. The integrated DLNM–LSTM framework was employed as methodologically complementary rather than redundant components, leveraging the distinctive strengths of each approach to address different analytical objectives within a unified investigation. Specifically, DLNM models characterize the non-linear exposure–lag–response relationships between meteorological factors and influenza risk, providing biologically interpretable insights into the temporal structure of weather-influenza associations and identifying meteorological factors with statistically and clinically significant effects on influenza dynamics. Building upon this DLNM-derived foundation, the LSTM network constructs a time-series forecasting tool within an identical covariate framework, evaluating predictive capability for influenza transmission trends. Beyond meteorological factors, our DLNM-LSTM framework systematically incorporated influenza positivity rates as the primary outcome variable, weekly detection volumes, mask-wearing stringency indices (indicator of NPIs during the COVID-19

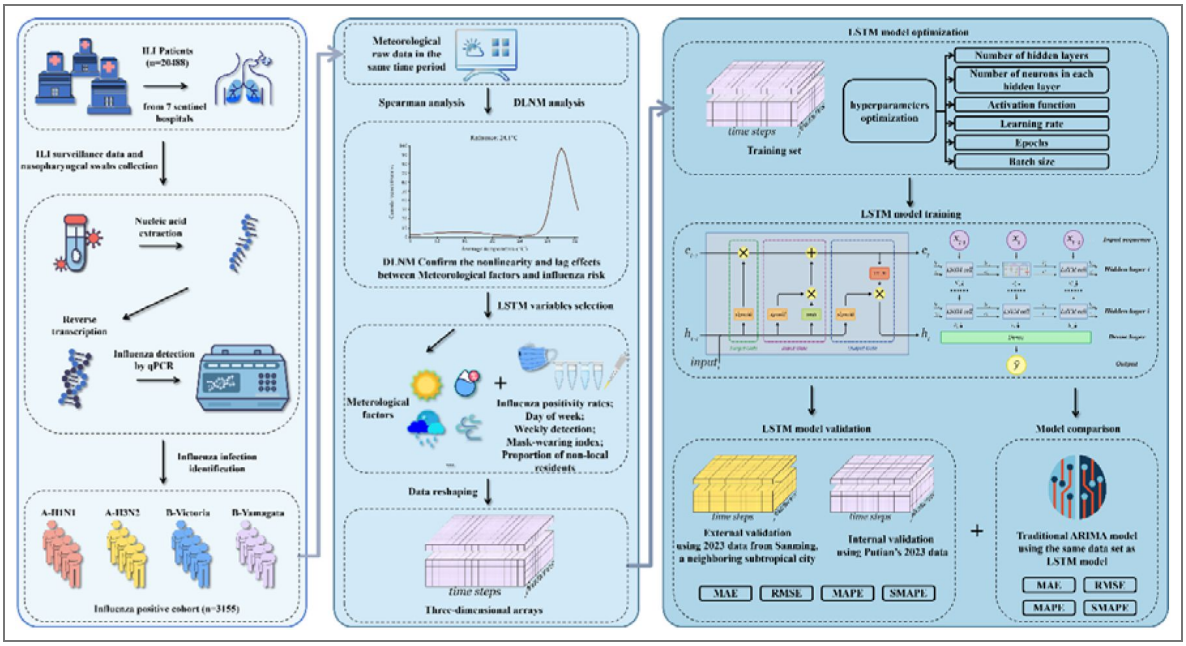

Figure 1. Schematic of the integrated analytical framework.

Sentinel surveillance data, daily meteorological records, and contextual covariates (weekly testing volume, weekday/weekend status, proportion of non-local residents, mask-wearing stringency indices) were jointly analysed using two complementary approaches: (i) distributed lag non-linear models (DLNM) to characterise non-linear, lagged exposure–response relationships between meteorological variables and subtype-specific influenza positivity, providing mechanistic interpretability; and (ii) a Bayesian-optimized Long Short-Term Memory (LSTM) neural network for short-term forecasting of daily positivity rates, benchmarked against a seasonal ARIMA model. The primary outcome variable for both analytical streams is the influenza positivity rate (laboratory-confirmed cases ÷ ILI specimens tested), selected to mitigate surveillance bias arising from temporal variation in testing intensity.

pandemic), day of the week (DOW, distinguishing weekdays from weekends), and non-local population proportion (defined as the ratio of the number of individuals whose reported residential district at the time of testing lies outside Putian city to the total number of tests) as covariates within both DLNM and LSTM modeling pipelines. This comprehensive covariate framework ensures that observed meteorological associations are estimated after controlling for surveillance intensity, public health intervention status, behavioral healthcare-seeking cycles, and population mobility patterns. The LSTM network was trained on data from Putian corresponding to the four categories described above, with the time period 2018-2022, and Putian's 2023 data serving as the internal validation set. To rigorously evaluate model transferability beyond the training context, data from Sanming city, a mountainous subtropical city exhibiting comparable climatic characteristics, influenza seasonality, and public health intervention frameworks to Putian, were acquired for the period January 1, 2023, to December 31, 2023, temporally aligned with the Putian internal validation set. This Sanming dataset constituted our external validation set, facilitating a geographically independent assessment of model generalization within subtropical Chinese environments. The predictive performance of the LSTM algorithm for influenza A/B prevalence in 2023 Putian data was benchmarked against a parallel multivariate ARIMA model with exogenous variables. To ensure a fair methodological comparison, this baseline model was supplied with the exact same meteorological and epidemiological covariate matrix as the LSTM.

Study cohort

The study systematically screened outpatients and emergency departments across seven influenza surveillance sentinel hospitals within Putian's prefecture, with an emphasis on high-incidence departments such as fever clinics, internal medicine, and pediatrics. The screening encompassed ILI cases from January 1, 2018, to December 31, 2023. According to earlier mentioned Diagnostic Criteria for Influenza, ILI cases are defined as acute onset with high fever (≥38°C) accompanied by respiratory symptoms such as cough or sore throat yet lacking a definitive causative diagnosis. Respiratory samples, together with onset dates, ages, genders, and residence information, were collected from all ILI patients. Unified diagnostic standards, laboratory testing methods, and quality control measures were employed throughout the research process to ensure the reliability and comparability of results.

After conducting nucleic acid screening for influenza viruses from the collected respiratory samples of ILI cases, the criteria for inclusion in this study prescribed the following: 1) the patient presented ILI symptoms; 2) the nucleic acid test for the influenza virus yielded a positive result; 3) the patient is a permanent resident of Putian city; 4) the duration of symptoms from onset to consultation did not exceed three days. No restrictions were placed on age or gender. The exclusion criteria were: 1) patients with clear alternative diagnoses, such as bacterial pneumonia or tuberculosis; 2) patients with unstable underlying conditions, such as acute exacerbations of chronic obstructive pulmonary disease (COPD); 3) patients with missing data. This focused the analysis on acute ILI presentations.

Sample collection and pathogen typing

Respiratory specimens (nasopharyngeal swabs) were collected from ILI patients during their initial visit, prior to treatment, and stored at 4°C in viral transport medium. All samples were delivered to the Putian CDC laboratory within 24 hours of collection. Nucleic acid extraction and one-step real-time fluorescent RT-PCR for influenza A/B subtyping (H1N1, H3N2, Victoria, and Yamagata lineages) were executed within 24 hours upon sample arrival. All assays utilized commercial diagnostic kits (supplied by Da'an Gene Co., Ltd., Guangzhou, China) and were processed in a Biosafety Level 2 (BSL-2) laboratory, strictly adhering to the manufacturer's instructions and China CDC's standardized protocols. Detailed laboratory procedures, including RNA integrity parameters and PCR amplification thresholds, are comprehensively documented in the Supplementary Information.

Meteorological and mask-wearing stringency data sources

Daily meteorological observation data were sourced from the authoritative public database Visual Crossing (*https://www.visualcrossing.com/*), which provides extensive historical time-series data across multiple meteorological variables on a global scale. We extracted meteorological variables for Putian city, including daily average temperature (°C), relative humidity (%), precipitation (mm), diurnal temperature range (°C), wind speed (km/h), wind direction, solar radiation (W/m^2), and the ultraviolet (UV) index, wherein the diurnal temperature range was defined as the difference between the highest and lowest temperatures of the day. All indicators were synchronized with the influenza data. To ensure data quality, we conducted missing value and anomaly checks on the raw datasets.

Early COVID-19 studies revealed a notable decline in positivity rates of seasonal respiratory viruses relative to pre-pandemic levels, likely due to containment measures such as masking and behavioral changes during surges (Groves et al. 2021; Xiao et al. 2022). To evaluate the potential impact of NPIs on influenza incidence, we obtained policies on face-covering stringency indices for China during the same period from the Our World in Data (OWID) database (*https://ourworldindata.org/coronavirus*) as a reference indicator for social prevention measures. This database categorizes stringency indices into four levels: 0 represents “No policy”; 1 represents “Recommended”; 2 represents “Required in some specified shared/public spaces outside the home with other people present, or some situations when social distancing not possible”; 3 represents “Required in all shared/public spaces outside the home with other people present or all situations when social distancing not possible”; 4 represents “Required outside the home at all times regardless of location or presence of other people”. While OWID stringency indices represent national-level policy, publicly accessible and standardized daily NPI datasets at the municipal level are currently unavailable in China. To ensure the spatial validity of these indices, co-authors from Putian CDC, who actively managed the local epidemic response, conducted a rigorous cross-validation. Our local public health experts confirmed that temporal fluctuations of OWID indices (ranging from Level 4 strict mandates in 2020 to Level 1 recommendations in 2023) exhibited high fidelity with the actual, on-the-ground enforcement of NPIs in both Putian and Sanming. Thus, OWID stringency indices serve as a highly reliable contextual proxy for local social contact restrictions. The datasets from these public databases, along with influenza surveillance data, are devoid of missing values, eliminating the need for imputation. For influenza and COVID-19 data, days with no reported cases were explicitly recorded as zero per the sentinel hospitals’ and government’s daily reporting protocols, respectively.

DLNM analysis

In the DLNM stage, we prioritized epidemiological interpretability by focusing on the isolated, non-linear, and lag-distributed main effects of individual meteorological variables. We deliberately avoided incorporating high-dimensional interaction cross-bases (e.g., temperature × humidity interactions) within the DLNM to prevent the “curse of dimensionality,” which typically induces severe multicollinearity, variance inflation, and loss of interpretability. Instead, the computational task of capturing complex, synergistic interactions among concurrent meteorological factors and epidemiological covariates was entirely delegated to the subsequent LSTM network. Separate DLNM models were constructed for influenza A and B incidence risks. To rigorously address the inherent confounding effects of “surveillance intensity bias”, where fluctuations in raw case counts may merely reflect transient surges in clinical testing capacity rather than true community transmission, the primary outcome metric for all DLNM modeling was mathematically defined as influenza positive rates, with meteorological factors serving and weekly detection volumes as independent variables. The formula for calculating the daily influenza positivity rate is as follows:

$$R(\%) = \frac{I}{N} \times 100\%$$

where *R* represents the daily influenza positivity rate, *I* is the number of daily influenza positive cases, and *N* denotes the total number of daily influenza tests performed. By adopting this WHO-recommended surveillance metric, our analytical framework explicitly standardizes the epidemiological denominator. This mathematical normalization effectively neutralizes the severe surveillance intensity bias caused by dramatic testing volume surges during the COVID-19 pandemic, ensuring that our models capture intrinsic viral transmissibility rather than artificial fluctuations in healthcare-seeking behavior or diagnostic capacity.

The maximum lag time was set to 15 days for all models, optimizing biological validity and analytical precision to capture typical influenza incubation and transmission cycles while minimizing model complexity and statistical noise from extended periods, aligning with previous research. A quasi-Poisson function was used to account for overdispersion in the incidence data. The model formula is as follows:

$$\log\left(E\left(y_{\text{time}}\right)\right) = \alpha + cb(x,k) + ns(\text{ time }, df)$$

where α is the intercept, *cd* is the cross-basis function of the meteorological factor *x*, which includes the exposure-response relationship ns(*x*,df_x), defined using natural splines and the lag-response relationship $g_{(k)}$ defined using polynomial functions, with , as the lag time; ns(time,df_{time}), is a natural spline function reflecting long-term trends, used to control for long-term trends and seasonality. The degrees of freedom df_x for the exposure-response relationship nsspline were selected from the range {2, 3, 4, 5}, while the degrees of freedom df_{time} per year for the long-term trend nsspline were selected from the range {6, 7, 8, 9, 10}. To determine the optimal *df*, including df_x for the natural spline modeling the exposure-response relationship and df_{time} per year for the natural spline modelling the long-term trend, the combinations of above ranges across selected meteorological factors were systematically evaluated using the quasi-Akaike information criterion (QAIC), with the combination yielding the smallest QAIC being selected. This indicates the best balance between goodness-of-fit and model complexity for our influenza-specific dataset. The QAIC formula is:

$$\text{QAIC } = -2\lambda + 2\varphi k$$

where λ represents the log-likelihood of the model, ϕ is the dispersion parameter, and *k* indicates the number of model parameters. When plotting prediction results, the median was selected as the reference value for its robustness against outliers, offering a consistent depiction of typical meteorological conditions in Putian and enabling the evaluation of bidirectional risk deviations, especially during extreme weather events. To further explore the impact of extreme weather conditions, we predicted and displayed the effects of extreme scenarios across the meteorological factors: except for precipitation extremes set at 5 mm and 25 mm and wind direction extremes at the most common eastern and rarest northwestern winds, all other meteorological factors were assessed at their 5^{th} and 95^{th} percentiles to estimate the lagged risk curves relative to the median. All results underwent seasonality and long-term trend corrections to eliminate the inherent temporal variation effects between meteorological factors and influenza incidence. Furthermore, to evaluate the robustness of our findings against testing intensity, we conducted a targeted sensitivity analysis by reconstructing the DLNM models without the “weekly detection volumes” covariate. We then compared the non-linear cumulative risks and extreme weather lag effects between these ablated models and the original fully adjusted models. This allowed us

to determine whether the identified climate-transmission associations were stable and biologically intrinsic, or merely artifacts of weather-correlated testing behaviors (Supplementary Information).

LSTM modeling and analysis

Building on the significant non-linear and lagged effects through DLNM analysis of real-world environmental exposures, this study further constructed multi-factor influenza A and B prediction LSTM networks. Leveraging a recurrent architecture with non-linear gating mechanisms, these networks are able to automatically capture and represent complex, high-dimensional interactions among meteorological variables without the need for manual pre-specification. This functional division of labor between the DLNM and LSTM models enhances predictive performance while preserving the interpretability and inferential stability established in the DLNM stage.

Data preprocessing

To respect the temporal dependence inherent in LSTM architectures and avoid data leakage, we adopted a strict chronological out-of-time (OOT) validation strategy. Time-series data from January 1, 2018, to December 31, 2022 (88.26% of the Putian dataset) were used for model training, with 10% reserved during Bayesian optimization as an internal validation subset for convergence monitoring and hyperparameter tuning only. Data from January 1, 2023, to December 31, 2023 (11.74%) were held out as a chronologically internal validation set for final performance evaluation, covering a complete annual cycle. External validation was conducted using concurrent 2023 data from Sanming city to assess spatial generalizability. To prevent distributional leakage, all normalization parameters were derived exclusively from the training set and consistently applied to the validation sets, with predictions subsequently transformed back to the original scale. First, we performed data normalization, a crucial step to ensure that training and test set data are compared on a unified scale. We normalized the training and test set data separately within the range [0, 1]. For the test set normalization, we used the maximum and minimum values from the training set as boundaries. This approach ensured consistency between the normalized test set data and the training set data. After the algorithm conducted predictions on the test set data, we performed denormalization to convert the predicted results back to the original data scale and rounded them to integers. These steps ensured that the final prediction results accurately and objectively reflected the LSTM's performance in real-world scenarios and provided reliable data for subsequent calculation of evaluation metrics.

The normalization formula is as follows:

$$y_{\text{normalization}} = \frac{y - \min_Y}{\max_Y - \min_Y}$$

The denormalization formula is as follows:

$$y_{\text{denormalization}} = \hat{y} \times (\max_Y - \min_Y) + \min_Y$$

where $y_{normalization}$ is the normalized result, y is the original data, max_y is the maximum value of the variable in the training set,min_y is the minimum value of the variable in the training set, $y1$ is the predicted value output by the network, and $y_{denormalization}$ is the denormalized result.

Finally, we used the above data while account for potential confounders by incorporating features such as DOW (1 for Monday–Friday, 0 for Saturday–Sunday), weekly detection volumes, non-local population proportion, and mask-wearing stringency indices, as input variables. The time step was set to 15 (i.e., using data from the previous 15 days to predict the number of influenza cases for the following day), and the two-dimensional data were organized into three-dimensional arrays to serve as input for the LSTM network.

Network construction

Our algorithm is a unidirectional LSTM, using Adam as the optimizer and Mean Square Error (MSE) as the loss function. The number of hidden layers was determined through hyperparameter optimization to ensure optimal network performance. Each LSTM layer consists of multiple LSTM units, including input gates, output gates, and forget gates. The basic framework of LSTM and the structure of LSTM units are shown in the LSTM model training section of Figure 1. Detailed principles and formulas are provided in the LSTM cell description section of Supplementary information and illustrated in Figure S1.

In the LSTM network, we used Hyperopt for hyperparameter tuning, with MSE established as the reference metric for evaluating and selecting optimal parameters. A detailed description of the Hyperopt Bayesian optimization strategy, including the specific search spaces for learning rates, network architecture (layers and neurons), and activation functions, along with their associated theoretical justifications, is provided in Supplementary Information.

LSTM performance evaluation

To comprehensively assess the predictive performance of the LSTM network, this study employed multiple commonly used error metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE) (Chicco, Warrens, and Jurman 2021).

MAE reflects the actual magnitude of the error between predicted and true values. A smaller MAE indicates more accurate network predictions. The calculation formula is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right|$$

RMSE is the arithmetic square root of the mean of the sum of squared errors. A smaller RMSE indicates better network prediction performance. The calculation formula is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right|^2}$$

A smaller MAPE indicates better network fitting. Since the denominator cannot be zero, this study only calculated MAPE for data where the true value was not zero. The calculation formula is:

$$\mathrm{MAPE} = \frac{1}{n}\times\sum_{i=1}^{n}\frac{\left|\hat{y}_i - y_i\right|}{\left|\hat{y}_i\right|}\times 100\%$$

A smaller SMAPE indicates better network fitting. The calculation formula is:

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

In the above formulas $\hat{y}_i$ is the denormalized predicted value, y_i is the observed value, and n is number of observations.

Additionally, we plotted the loss function curves for the network on the training and validation sets to monitor LSTM convergence and the risk of overfitting. By comparing the trends and differences between the two curves, we could intuitively judge the training quality and generalization ability of the LSTM. Furthermore, the Bayesian-optimized LSTM architecture was directly applied, without re-training or hyperparameter adjustment, to both the Putian internal

test set and the Sanming external validation set. This unified modeling framework ensures that performance differences between the two evaluation contexts primarily reflect geographic transferability rather than model re-optimization.

To further explain the prediction logic of the LSTM, this study employed the SHapley Additive exPlanations (SHAP) method for interpretability analysis. Based on Shapley values from game theory, SHAP can quantify the contribution of each feature to the LSTM's predicted values and generate intuitive feature importance rankings and dependency plots. Positive values indicate that the feature increased the predicted result, while negative values indicate a decrease in the predicted result, with the absolute value representing the intensity of the influence. SHAP values were computed by treating each meteorological variable's lagged sequence as a unified input feature, preserving the temporal dependencies within each sequence rather than decomposing lags into independent features. This approach not only aligns with the sequential nature of our LSTM inputs but also supports our focus on population-level trends rather than individual-level causal effects.

To compare the predictive performance between the LSTM and traditional approaches, we concurrently developed a ARIMA model, employing the same training and validation datasets (detailed methodological description in Supplementary Information). To assess the contribution of pandemic-related confounding variables, we performed targeted sensitivity analyses on the LSTM networks. Specifically, we sequentially removed covariates of mask-wearing stringency indices and weekly detection volumes from the input features while maintaining identical Bayesian-optimized hyperparameters. The performance of these ablated models was evaluated using MAE, RMSE, MAPE, and SMAPE, allowing us to quantify the exact necessity of incorporating testing and behavioral covariates in forecasting models during periods of epidemiological non-stationarity.

Statistical analysis

Statistical analyses were primarily conducted using R software version 4.3.1 and Python version 3.8.3. DLNM modeling was implemented using the DLNM package in R. LSTM construction used Python 3.8.3, with the main program packages including tensorflow 2.2.0, keras 2.3.1, and shap 0.44.1. The ARIMA model was developed using the forecast package in R. All statistical tests were two-sided, with $P < 0.05$ considered statistically significant.

Results

Descriptive epidemiological characteristics of influenza

Between 2018 and 2023, influenza surveillance conducted at seven sentinel hospitals in Putian city documented a total of 20,488 ILI cases, with the majority being influenza-negative (17,333 cases, 84.60%). Of these, 3155 cases were confirmed positive for influenza through PCR testing, representing a positivity rate of 15.40%. By viral type, 1685 cases were attributed to influenza A (8.23%), and 1470 cases were influenza B (7.17%). Further subtype analysis of influenza A revealed 790 cases of H1N1 (46.88%) and 895 cases of H3N2 (53.12%), while for influenza B, 1270 cases were of the Victoria lineage (86.39%) and 200 cases were of the Yamagata lineage (13.61%). Notably, following the onset of the COVID-19 pandemic in the spring of 2020, there was a significant reduction in the incidence of influenza. The annual incidence rates for influenza in 2020 and 2021 were 4.54 per 100,000 and 5.59 per 100,000, respectively, markedly lower than the pre-pandemic rates observed in 2018 (16.01 per 100,000) and 2019 (18.74 per 100,000), suggesting that public health measures implemented during the COVID-19 pandemic, such as mask-wearing and social distancing, were highly effective in decreasing the transmission of respiratory infections, including influenza.

In terms of seasonal distribution, surveillance data (Figure 2 A) revealed a notable dynamic interplay between influenza A and B viruses over the study period. In 2018, influenza A H1N1 was the dominant strain. However, by 2019, cases of the influenza B Victoria lineage gradually increased, surpassing H1N1 during the summer to become the predominant strain. In 2020, there

was a sharp decline in influenza cases, with influenza A H3N2 emerging as the leading strain. The dominance of influenza B Victoria returned in 2021, only to be overtaken by a resurgence of influenza A H3N2 in the summer of 2022. By the winter of 2023, both influenza A H3N2 and influenza B Victoria exhibited concurrent surges, highlighting a dual epidemic. Interestingly, contrary to the conventional understanding that influenza A typically drives larger epidemics, our findings show that influenza B accounted for 52.18%, 100%, and 64.76% of total influenza cases in 2019, 2021, and 2022, respectively.

In the demographic analysis (Figure 2 B), among the 3155 laboratory-confirmed influenza cases, males (1811 cases; 57.40%) outnumbered females (1344 cases; 42.60%), with a male-to-female ratio of 1.35:1. Significant gender differences were observed across age groups ($\chi^2 = 37.36$, $P < 0.001$), with males constituting 60.68% of cases in individuals under 18 years, whereas females were more prevalent among adults (51.53%). A statistically significant difference was noted in the gender distribution between influenza A and B ($\chi^2 = 6.64$, $P = 0.01$), with males comprising a higher proportion (59.86%) of influenza B cases. No statistically significant gender differences were found in the distribution of influenza subtypes ($\chi^2 = 2.43$, $P = 0.49$). Additionally, children and adolescents under 18 years of age were identified as the most at-risk population, comprising 73.12% of all cases. Influenza B was more prevalent in this age group (50.33% vs. 49.67%), while influenza A predominated in the adult population (63.56% vs. 36.44%). These findings suggest that susceptibility to different influenza virus types may vary across age groups.

Overall, our findings suggest that in subtropical regions, the epidemic intensity of Influenza B may rival or even exceed that of Influenza A. Furthermore, the widespread implementation of non-pharmaceutical interventions (e.g., mask-wearing) during the COVID-19 pandemic significantly curtailed the transmission of respiratory viruses, including influenza. Continuous monitoring of Influenza A (H1N1, H3N2) and Influenza B Victoria strains is essential, alongside heightened surveillance and preventive measures targeting children and adolescents, particularly in densely populated settings such as schools. These insights are crucial for informing vaccine strain selection, vaccination strategies, and targeted public health interventions.

Temporal trends in meteorological factors and influenza incidence

Between 2018 and 2023, influenza incidence in Putian city exhibited distinct seasonal patterns. The majority of cases occurred between December and April of the following year, accounting for over 60% of the annual cases, with a secondary peak observed from May to September. Concurrently, meteorological factors demonstrated clear cyclical variations (Figure 3), with detailed descriptive statistics provided in Supplementary Table S1. During the summer months (July–August), the average daily temperature peaked above 32°C, while the winter months (December to February) experienced the lowest temperatures, dropping below 10°C (Figure 3 A). This period was also marked by a significant increase in diurnal temperature variation, with the maximum daily range exceeding 13°C (Figure 3 E). The annual average relative humidity was approximately 76.10%, with the highest levels observed during the monsoon season from April to June, reaching or exceeding 97% (Figure 3 B). Precipitation showed a pronounced seasonal distribution, peaking from June to September, while the period from November to February was characterized by markedly reduced rainfall (Figure 3 C). Solar radiation (Figure 3 D) and UV index (Figure 3 F) both reached their annual maxima between July and September. High wind speeds were frequently observed in Putian city from July to September during the study period, which may correlate with the increased frequency of typhoons during this period (Figure 3 G). Throughout the year, the easterly wind direction was predominant, with infections in every month (Figure 3 H).

A comparative analysis of the temporal patterns of influenza cases and meteorological factors revealed several noteworthy correlations. Firstly, peaks in influenza A and B cases were predominantly observed during January to March, characterized by lower average temperatures and larger diurnal temperature ranges (Figures 3 A and E). Interestingly, influenza B occasionally exhibited summer outbreaks, indicating a potentially greater adaptability to higher temperatures. Secondly, extreme rainfall events and periods of high humidity were often

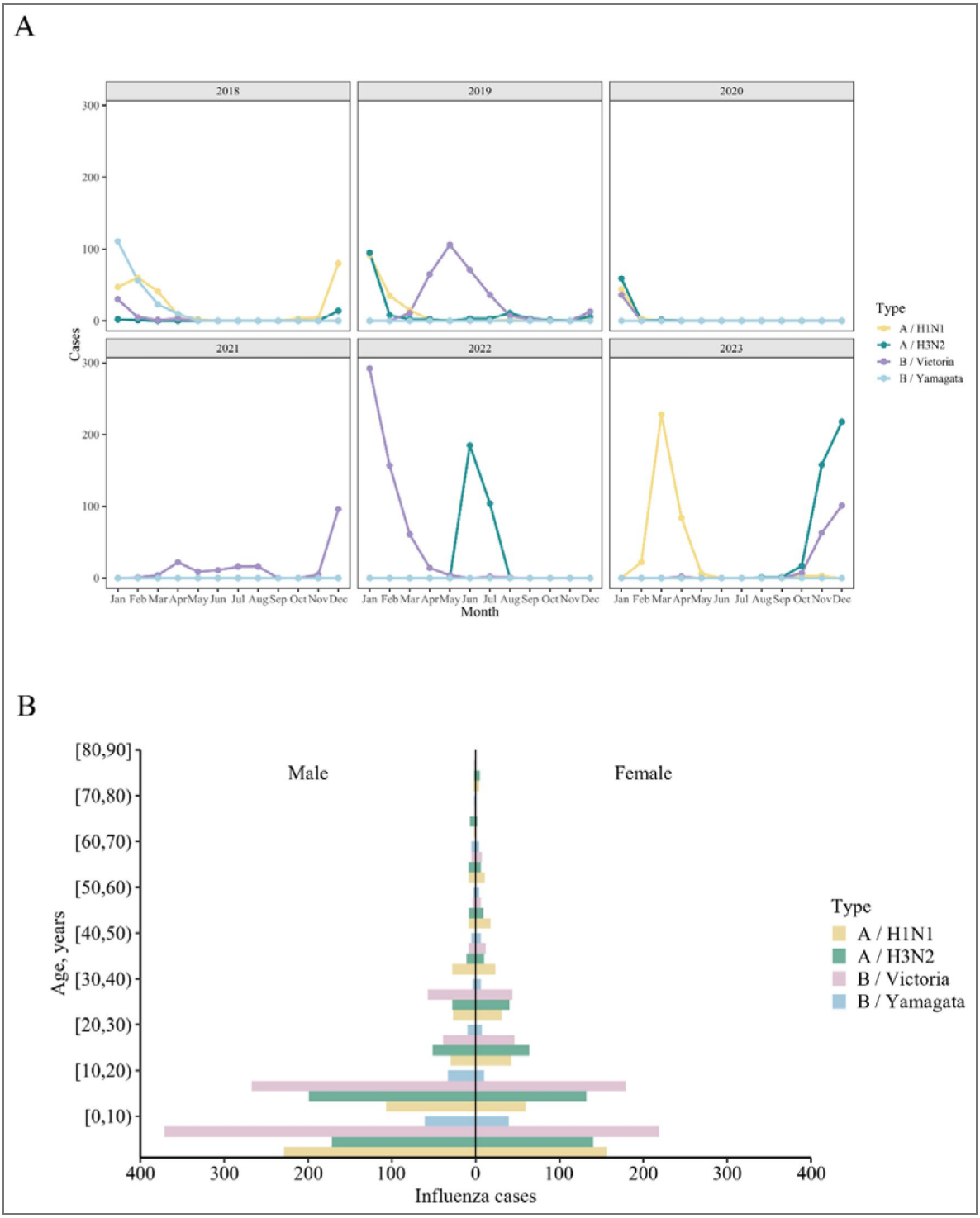

Figure 2. Temporal trends and demographic distribution of confirmed influenza cases in Putian city (2018 to 2023).

This figure presents the temporal and demographic characteristics of influenza cases in Putian City over the period from 2018 to 2023. **(A)** Line plots illustrating the monthly case counts of various influenza subtypes (Influenza A: H1N1, H3N2; Influenza B: Victoria, Yamagata) across the six-year study period. The transition from historical heatmaps to precise line plots enhances the visualization of distinct epidemiological surges, notably capturing the profound interruption of typical seasonal transmission during the peak COVID-19 pandemic restrictions (2020–2021) and the subsequent viral rebound. **(B)** A population pyramid detailing the age and sex distribution of the 3,155 laboratory-confirmed influenza cases. The visualization bifurcates cases by gender (males on the left, females on the right) across age intervals, underscoring the differential susceptibility of specific demographic cohorts, particularly highlighting the heightened vulnerability of children (0-10 years) and the elderly (60 years and above).

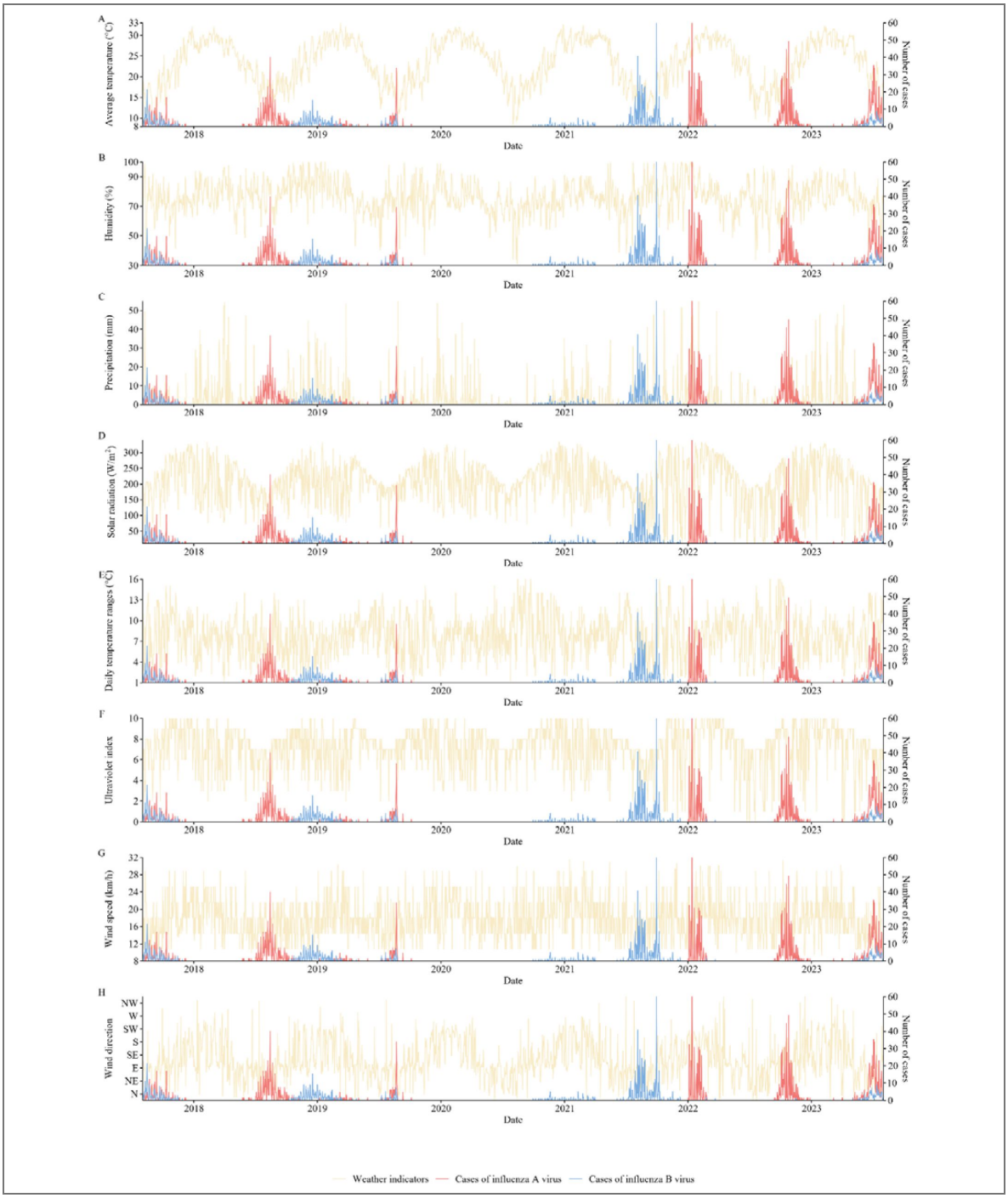

Figure 3. Time-series dynamics of meteorological factors and absolute influenza incidence in Putian city from 2018 to 2023.

This figure vertically aligns temporal fluctuations of influenza-related and key meteorological variables, January 2018-December 2023. From top to bottom: **(A)** average temperature, **(B)** humidity, **(C)** precipitation, **(D)** solar radiation, **(E)** daily temperature ranges, **(F)** ultraviolet index, **(G)** wind speed, and **(H)** wind direction, plotted continuously over the five-year surveillance period. In each panel, the left y-axis corresponds to the observed values of the respective meteorological factor (yellow lines), while the right y-axis delineates the absolute number of confirmed cases for Influenza A (red lines) and Influenza B (blue lines). These descriptive plots visually contextualize the raw seasonal correlations and climatic extremes preceding major influenza outbreaks. Note the substantial suppression of influenza activity during 2020–2022 (pandemic intervention period) and the pronounced post-restriction rebound during 2023, particularly for influenza A, which exceeded pre-pandemic peak levels.

temporally aligned with short-term surges in influenza A and B cases (Figures 3 B and C). For instance, the peak in influenza B cases in June 2022 coincided with an episode of extreme rainfall. Moreover, the majority of influenza outbreaks occurred during periods of reduced solar radiation and lower UV indices (Figures 3 D and F). Variations in wind speed were not significantly correlated with influenza incidence in our analysis (Figure 3 G). During periods of easterly winds, there was a noticeable decrease in the incidence of influenza A and B cases in Putian city (Figure 3 H). Our findings indicate that the incidence of influenza in the subtropical city of Putian was closely associated with various meteorological conditions.

DLNM analysis for the cumulative and lagged effects of multiple meteorological factors on influenza incidence risks

Prior to analytical modeling, we assessed the pairwise associations and potential collinearity among all meteorological variables using a comprehensive correlation heatmap and scatterplot matrix (Supplementary Figure S2). Notably, certain variables, such as solar radiation and the UV index, exhibited high positive correlations. To avoid coefficient instability typically caused by multicollinearity in regression models, all DLNM analyses adopted a univariate approach for meteorological factors, sequentially evaluating the non-linear and lagged effects of individual meteorological predictors while adjusting for time trends and including other fixed covariates. Conversely, all variables were retained during the LSTM forecasting phase, as the non-linear gating architecture of recurrent neural networks inherently exhibits robust regularization against potential collinearity among input features.

To elucidate the non-linear associations between meteorological factors and influenza transmission dynamics, DLNM models were constructed using influenza positivity rates as the outcome variable. For each meteorological variable, the median exposure level was used as the reference. For each influenza-specific DLNM model, the combination of df_x for the exposure-response spline and df_{time} for the long-term trend spline that yielded the minimum QAIC was selected as the optimal configuration and applied to the final model. Detailed QAIC values for all tested combinations are provided in Supplementary Table S2. Overall, the associations between the eight evaluated meteorological factors and the cumulative risk of influenza positivity exhibited pronounced non-linear and subtype-specific characteristics over the 15-day lag period (Figure 4 A and B). Influenza A was primarily associated with lower temperature, reduced diurnal temperature range, and intermediate UV levels, whereas influenza B showed broader meteorological sensitivity, with increased risks observed under lower temperature, higher humidity, lower solar radiation, lower UV index, and specific wind directions.

Regarding temperature effects, using 24.1°C as the reference, the cumulative risk for influenza A showed a non-linear inverse-U pattern, peaking at 15°C with a relative risk (RR) of 3.01 (95% CI: 1.03-8.78) and gradually decreasing across the 15–24°C range. In contrast, the cumulative risk for influenza B peaked at a much lower temperature of 8°C (RR = 26.12, 95% CI: 7.79–87.61), with the risk steadily declining between 8°C and 24°C.

Humidity demonstrated divergent effects between the two subtypes. While relative humidity showed no significant association with influenza A risk (using 75.8% as a reference), the cumulative risk for influenza B increased gradually within the high-humidity range of 76% to 91%, reaching its peak at 91% (RR = 1.52, 95% CI: 1.00–2.29). Similarly, precipitation showed no significant impact on influenza A. However, for influenza B, precipitation within the 0–10 mm range exhibited a rise-then-fall risk pattern, peaking at 5 mm (RR = 4.25, 95% CI: 2.11–8.58). These findings suggest that influenza A is predominantly driven by specific temperature bands, whereas influenza B transmission is more closely linked to cool-wet meteorological regimes.

Solar radiation and the UV index also manifested distinct subtype affinities. For influenza A, neither low (75 W/m²) nor high (305 W/m²) solar radiation yielded significant cumulative risks. In contrast, low solar radiation (75 W/m²) posed a significant risk for influenza B, peaking at 9.6 days

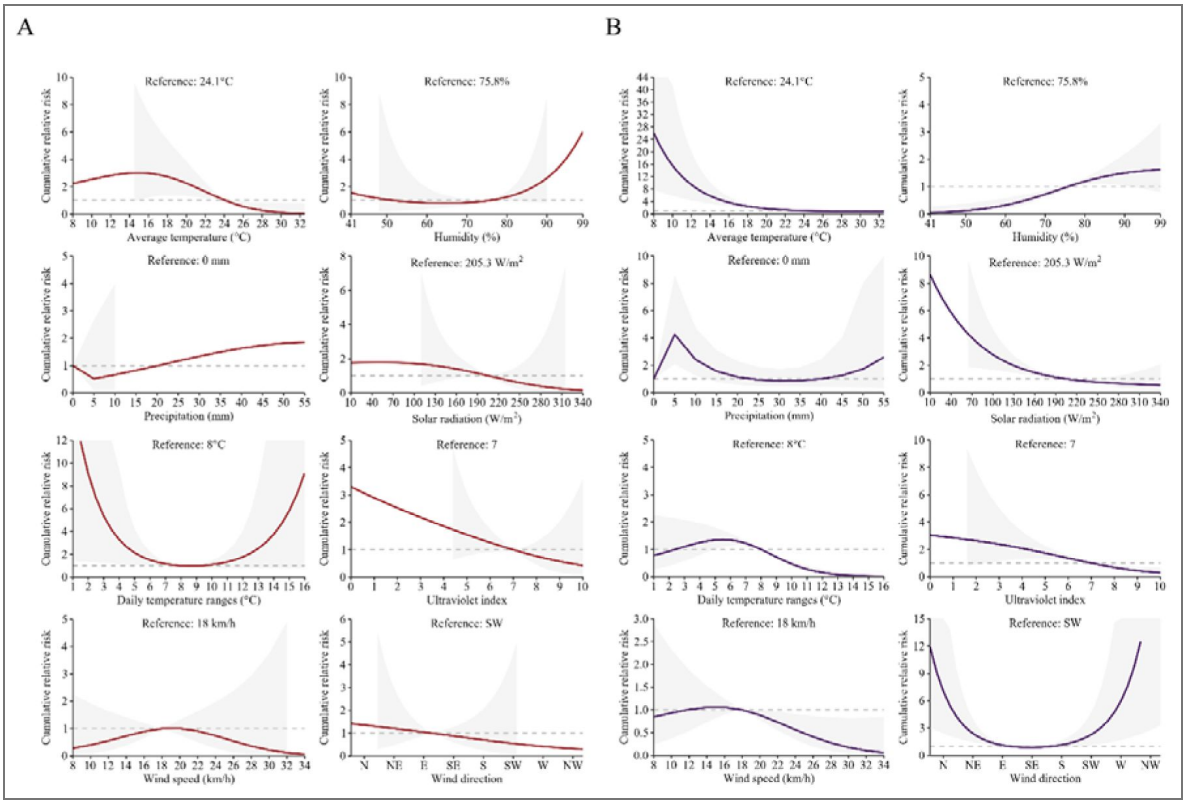

Figure 4. Cumulative risk of meteorological factors on Influenza A and B positivity rates over a 15-day Lag period.

This figure illustrates the non-linear, cumulative exposure-response relationships between eight individual meteorological drivers and the 15-day cumulative relative risk of influenza A and B positivity, estimated via Distributed Lag Non-linear Models (DLNM). Models were adjusted for long-term trends, weekly detection volumes, and other fixed covariates. Panels in **(A)** depict the non-linear cumulative risks for Influenza A, while panels in **(B)** shows the corresponding risks for Influenza B. In each subplot, the x-axis represents the observed continuous range of a specific meteorological factor, and the y-axis indicates the cumulative relative risk compared to a predefined median reference value (e.g., 24.1°C for average temperature, 75.8% for humidity). The solid line reflects the estimated non-linear trend of risk variation, and the shaded regions denote the 95% confidence intervals (CIs).

(RR = 1.16, 95% CI: 1.08–1.23). For the UV index (reference = 7), influenza A only showed significant risk exactly at an index of 7 (RR = 1.00, 95% CI: 1.00–1.00), whereas influenza B risk progressively decreased within the UV index range of 2.6 to 7, peaking at 2.6 (RR = 2.49, 95% CI: 1.04–5.96).

Diurnal temperature range (DTR) was associated with both influenza subtypes, although the estimates for influenza A were imprecise. Relative to 8°C, the cumulative risk of influenza A was highest at 1°C (RR = 15.69, 95% CI: 1.42–173.00), while influenza B risk peaked at 5.5°C (RR = 1.36, 95% CI: 1.09–1.69). Wind speed was not significantly associated with either influenza subtype. Wind direction was not associated with influenza A, but northerly, easterly, and westerly winds were associated with elevated influenza B risk, with the highest estimate observed for northerly winds (RR = 27.33, 95% CI: 3.33–224.41).

Lagged effects of extreme meteorological conditions on influenza positivity risks

We further investigated the dynamic temporal variations in influenza positivity risks under extreme meteorological conditions (evaluating the 5^{th} and 95^{th} percentiles of each variable) to characterize type-specific vulnerability windows (Figures 5 A and B). Low temperature was associated with short-term increases in influenza A risk, with significant effects between lag 0.2 and 1 day and a peak at lag 0.2 days (RR = 1.55, 95% CI: 1.01–2.39). High temperature was not associated with influenza A. For influenza B, low temperature showed both immediate and delayed associations, with elevated risks at lag 0–5.6 days and 9.2–10.2 days. The strongest effect occurred at lag 0 days (RR = 1.56, 95% CI: 1.26–1.93). High temperature was also associated with influenza B risk at lag 4.2–7.2 days, peaking at lag 5.8 days (RR = 1.13, 95% CI: 1.04–1.23).

Under extreme humidity, neither very low (58%) nor very high (94%) humidity generated significant lagged risks for influenza A. However, for influenza B, low humidity posed a rapid risk (1.2–3 days, peak RR = 1.11 at day 2), while high humidity resulted in a delayed risk (5–9.2 days, peak RR = 1.08 at day 7.2). For precipitation, heavy rainfall (25 mm) uniquely impacted influenza B, inducing a lagged risk between 7.8 and 11 days (peak RR = 1.12 at day 9.6), with no significant effects observed for influenza A.

Low solar radiation and low UV index were consistently associated with delayed increases in influenza B risk. Low solar radiation increased influenza B risk from lag 4.6 to 12.6 days, peaking at lag 9.6 days (RR = 1.16, 95% CI: 1.08–1.23). Low UV index was associated with elevated influenza B risk from lag 5 to 11.8 days, peaking at lag 9.2 days (RR = 1.12, 95% CI: 1.05–1.20). No significant lagged associations were observed between these radiation-related variables and influenza A.

Extreme DTR showed limited but subtype-specific lagged effects. Low DTR was associated with influenza A risk at lag 0.2–2 days and 11–12.8 days, with peak estimates at lag 0.2 days (RR = 1.31, 95% CI: 1.02–1.66) and lag 12.2 days (RR = 1.19, 95% CI: 1.01–1.40). For influenza B, low DTR was associated with increased risk at lag 10.2–12.2 days, peaking at lag 11.8 days (RR = 1.10, 95% CI: 1.00–1.20). High DTR showed no significant association with either subtype. Wind-related effects were observed primarily for influenza B. Southwesterly winds were associated with increased influenza B risk from lag 0.4 to 10.4 days, peaking at lag 2.4 days (RR = 1.30, 95% CI: 1.17–1.44), while easterly winds were associated with risk from lag 5 to 9.8 days, peaking at lag 7.4 days (RR = 1.16, 95% CI: 1.07–1.26). Wind speed was not significantly associated with either influenza A or B.

Crucially, sensitivity analyses indicated that after removing the “weekly detection volumes” covariate, the reconstructed DLNM models remained structurally highly consistent with the original fully-adjusted models. This robust concordance confirms that the identified meteorological drivers represent intrinsic environmental triggers of viral transmission, independent of fluctuating surveillance intensity. Detailed methodological descriptions and results of the sensitivity analysis are presented in Supplementary information.

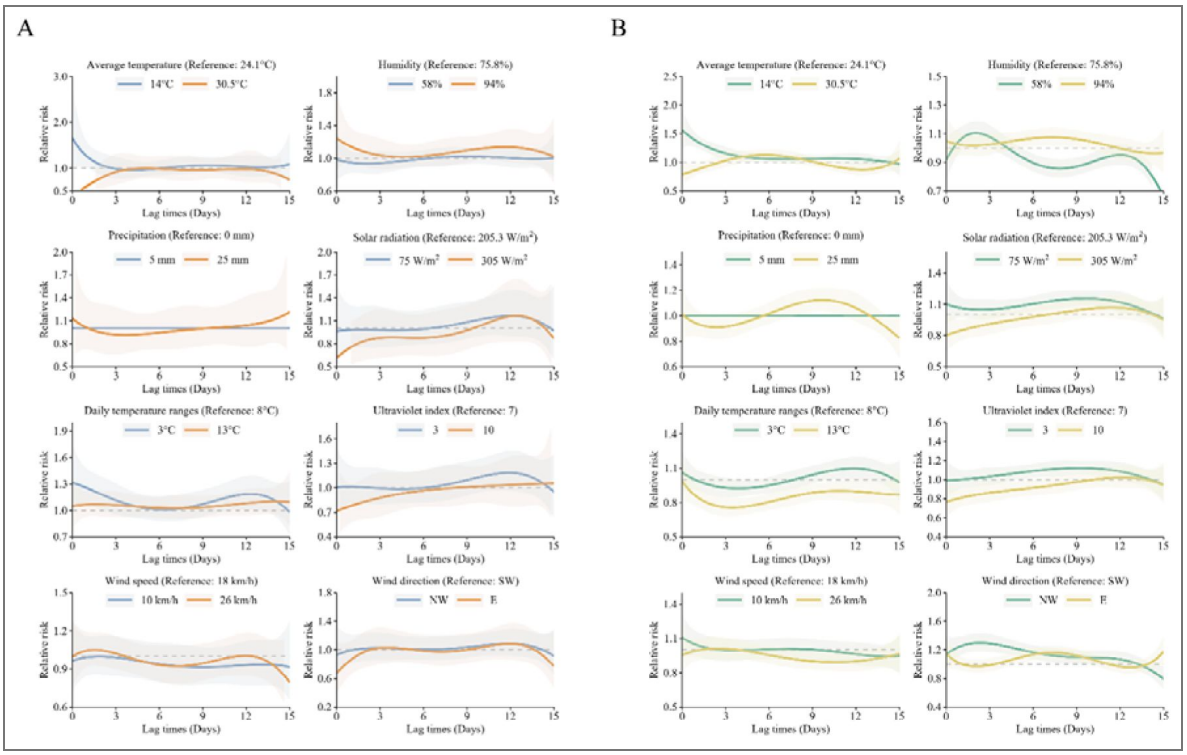

Figure 5. Lagged temporal effects of extreme meteorological conditions on Influenza A and B positivity rates.

This figure visualizes the distribution of relative risks over a 15-day lag period following exposure to extreme meteorological conditions, derived from Distributed Lag Non-linear Model (DLNM). The analysis compares the temporal lag effects at the extreme high (95th percentile) and extreme low (5th percentile) values of selected weather variables for Influenza A (**A)** and Influenza B (**B**). In each subplot, the x-axis represents the specific lag days (from day 0 to 15), and the y-axis shows the relative risk compared to the median reference condition. Solid orange lines (for Influenza A) and yellow lines (for Influenza B) represent extreme high-value conditions (e.g., extreme heat), whereas solid blue lines (for Influenza A) and green lines (for Influenza B) represent extreme low-value conditions (e.g., extreme cold). The shaded areas surrounding the curves represent the 95% confidence intervals. These lag-response curves elucidate the dynamic delay between specific extreme weather events and the subsequent peaks in influenza test positivity.

Influenza prediction using LSTM network

Building upon the previous analysis of the impact of meteorological factors on influenza dynamics, this study further aimed to construct a multifactorial influenza prediction network utilizing LSTM. To optimize the LSTM network, we employed Bayesian optimization through the Hyperopt framework, which allowed us to fine-tune key hyperparameters. This optimization process led to the identification of the optimal configurations for influenza A and B prediction algorithm (Table 1).

For influenza A, the optimal LSTM network consisted of one hidden layers with 256 neurons, a learning rate of 0.0001, and a batch size of 8. In contrast, the influenza B LSTM network had two hidden layers with 512 and 256 neurons, a learning rate of 0.0001, and a batch size of 8. Leveraging these optimal hyperparameters, alongside a combination of meteorological data, weekly detection volumes, DOW, non-local population proportion, and mask-wearing stringency indices, we trained the LSTM network using historical data from 2018 to 2022. A 10% of the data were randomly selected as the validation set, while data from 2023 were subsequently used to test the network's performance. The loss curves plotted during training revealed the convergence behavior of the LSTM. As shown in Figures 6 C and D, the training loss and validation loss for both the influenza A and B LSTM networks decreased progressively with increasing epochs, initially displaying a steep decline followed by stabilization. This pattern indicates that the algorithms effectively learned the characteristics of the data and converged to a stable solution. Furthermore, the close alignment between training and validation loss curves suggests that both LSTM networks for influenza A and B performed well and exhibited minimal overfitting, demonstrating strong learning capabilities and generalizability.

Following adequate training, the LSTM networks were employed to predict influenza A and B positivity rates in Putian city for the year 2023. The results revealed a strong match between the predicted and observed values (Figures 6 A and B). The LSTM networks accurately captured both the timing and magnitude of these non-linear epidemic surges, including the two outbreak peaks of influenza A during February-March and November-December of 2023, as well as the peak of influenza B in November-December, demonstrating good predictive performance. The predictive performance of the LSTM networks was quantitatively assessed using metrics such as MAE, RMSE, MAPE, and SMAPE. For influenza A, the MAE was 0.009, RMSE was 0.035, MAPE was 0.158, and SMAPE was 0.521; for influenza B, the MAE was 0.002, RMSE was 0.011, MAPE was 0.17, and SMAPE was 0.484 (Table 2). These findings underscore the precision of the LSTM in characterizing and predicting influenza trends.

To quantify the contributions of epidemiological and behavioral covariates to the network's handling of pandemic non-stationarity, we conducted a covariate-ablation sensitivity analysis (Table 2). The LSTM networks were retrained sequentially by removing mask-wearing stringency indices and weekly detection volumes, while maintaining identical Bayesian-optimized hyperparameters. For influenza A, the exclusion of mask-wearing stringency indices resulted in a pronounced degradation of predictive performance compared to the full model: MAE, MAPE, and SMAPE increased by 33.3%, 56.3%, and 62.4%, respectively. The removal of weekly detection volumes similarly degraded performance, increasing MAE by 22.2% and SMAPE by 51.8%. This systemic shift in absolute errors (MAE) and the amplification of relative errors (SMAPE) clearly indicate that LSTM actively leverages mask mandates and testing intensities to capture the relative trends of influenza A. For influenza B, the ablation of covariates revealed an even higher dependency on the surveillance context. Removing weekly detection volumes proved highly detrimental, triggering a 100% surge in MAE, an 18.2% increase in RMSE, and a drastic 199.6% spike in SMAPE. This confirms that testing volumes act as a critical constraint variable for forecasting influenza B. Omitting mask-wearing indices similarly increased MAE by 50% and SMAPE by 77.1%.

Notably, a divergence was observed in the MAPE metric for influenza B, which paradoxically decreased (by 20.1% and 29.4%) upon the removal of these two covariates. This phenomenon is a well-documented mathematical artifact of the MAPE calculation, which utilizes the true observed

Hyperparameter category	The number of hyperparameter for predicting influenza A positivity rate	The number of hyperparameter for predicting influenza B positivity rate
Layers	1	2
Neurons in layer 1	256	512
Neurons in layer 2	-	256
Neurons in layer 3	-	-
Activation function	sigmoid	softplus
Batch size	8	8
Learning rate	0.0001	0.0001
Epochs	587	931

LSTM: Long short-term memory. Hyperparameters were optimized through grid search. This table presents the optimal LSTM network structure (number of layers and neurons), activation function, batch size, learning rate, and number of epochs for predicting both influenza A and B infections. The number of neurons refers to the count of LSTM cells in each layer; detailed LSTM architecture is provided in Supplementary Information Additional file 1.

Table 1. Optimal LSTM hyperparameters for predicting influenza A and B positivity rates.

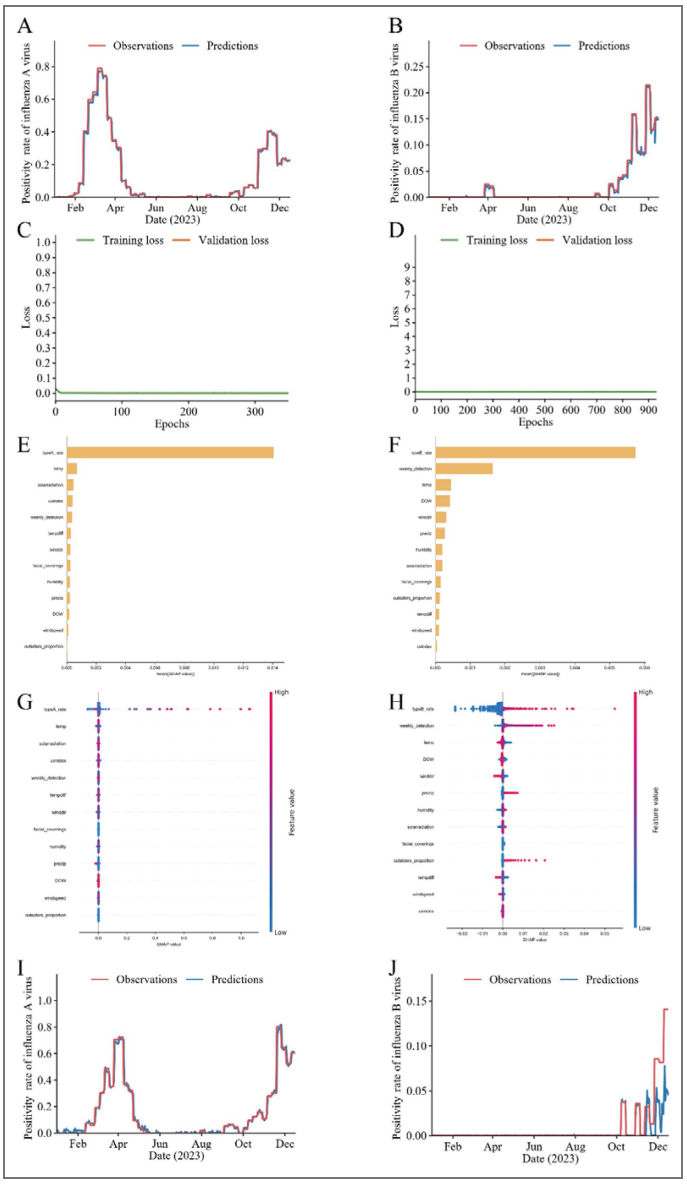

Figure 6. High-fidelity forecasting of the Influenza Positive Rate using the LSTM network and SHAP interpretability.

This comprehensive panel visualizes the predictive performance and feature importance of the LSTM model configured to forecast the influenza positivity rate. **(C)** and **(D)** Training loss curves: These panels show the variation in loss values during the training of the LSTM network. The x-axis represents the number of training epochs, while the y-axis indicates the loss values. The smooth decline and stabilization of the loss curves indicate effective learning by the network, with no significant signs of overfitting. This is a crucial measure of the network's stability and effectiveness throughout the training process. **(A)** and **(B)** Internal Out-of-Time validation (2023) comparing the LSTM-predicted positive rate (blue lines) against the observed positive rate (red lines) for Influenza A and B. The x-axis represents the date, and the y-axis reflects the positivity rates. This comparison visually demonstrates the network adeptly captures the non-linear viral rebound dynamics. **(E)** and **(F)** SHAP (SHapley Additive exPlanations) feature importance rankings. The bar charts quantify the mean absolute SHAP values, identifying historical subtype positivity rates (typeA_rate / typeB_rate), testing volume (weekly_detection), average temperature (temp), and mask-wearing stringency indices (facial_coverings) as paramount predictive drivers. The x-axis denotes the mean SHAP value, while the y-axis lists the variables. **(G)** and **(H)** SHAP summary plots illustrating the distribution of impacts for each feature on the model's output, where color indicates feature value (red = high, blue = low). This graph provides a visual understanding of how and to what extent different features influence the influenza predictions, offering critical insights into the network's decision-making process. (I) and (J) External spatial validation utilizing 2023 surveillance data from the neighboring city of Sanming, confirming the LSTM's robust generalizability in forecasting independent epidemic curves. The date is plotted on the x-axis and the positivity rates are plotted on the y-axis.

Performance evaluation metric	Predicting Influenza A Positivity Rate			Predicting Influenza B Positivity Rate		
	Full Model (All variables included)	Mask-wearing index Removed	Weekly detection volumes Removed	Full Model (All variables included)	Mask-wearing index Removed	Weekly detection volumes Removed
MAE	0.009	0.012	0.011	0.002	0.003	0.004
RMSE	0.035	0.035	0.034	0.011	0.012	0.013
MAPE	0.158	0.247	0.203	0.170	0.135	0.120
SMAPE	0.521	0.846	0.791	0.484	0.857	1.450

Note: MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) represent absolute predictive deviations. MAPE (Mean Absolute Percentage Error) and SMAPE (Symmetric Mean Absolute Percentage Error) assess relative accuracy. The full model incorporates meteorological variables alongside epidemiological covariates (mask-wearing stringency indices, weekly detection volumes, DOW, and migrant population proportion). Ablated models demonstrate the degradation in predictive accuracy upon the targeted removal of specific covariates, highlighting their necessity in adjusting for epidemiological non-stationarity.

Table 2. Sensitivity analysis of the LSTM forecasting network evaluating the impact of epidemiological covariate ablation on predictive performance. Note: MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) represent absolute

value as the denominator. Given the extremely low baseline positivity rate of influenza B during certain periods, infinitesimal absolute deviations are asymmetrically magnified into massive percentage errors, causing MAPE to become highly unstable and misleading. In contrast, SMAPE, which utilizes the mean of the predicted and actual values in the denominator, effectively circumvents this asymmetric "low base-rate effect". The profound deterioration of SMAPE across all ablated models corroborates the positive, stabilizing contribution of NPI and testing-volume covariates. These findings validate that our unified LSTM architecture does not merely rely on meteorological inputs, but effectively contextualizes epidemiological regime shifts to optimize forecasting precision.

To rigorously evaluate the spatial generalizability of the LSTM framework, we conducted an external validation utilizing concurrent 2023 surveillance and meteorological data from Sanming city. The network demonstrated robust transferability, achieving an MAE of 0.015, RMSE of 0.044, MAPE of 0.176, and SMAPE of 0.674 for influenza A; and an MAE of 0.005, RMSE of 0.018, MAPE of 0.487, and SMAPE of 1.792 for influenza B. As illustrated in Figures 6 I and J, the predicted trajectories successfully mirrored the actual observed incidence patterns in the new region, confirming the framework's operational stability beyond its internal validation domain. To benchmark the predictive value added by the deep learning architecture, we constructed multivariate ARIMA models using identical training and validation positivity-rate datasets, inclusive of all contextual covariates. The LSTM model demonstrated decisive superiority over this covariate-adjusted ARIMA baseline across all evaluated metrics. For influenza A, the ARIMA model yielded a substantially higher MAE of 0.136 and SMAPE of 1.212; for influenza B, it produced an MAE of 0.049 and SMAPE of 0.810. Beyond the global error metrics, a visual comparison of the forecasting trajectories (Supplementary Figure S5) reveals critical behavioral disparities. For influenza A, the linear ARIMA model generated conservatively smoothed forecasts that entirely failed to capture the sudden, explosive non-linear peaks of the 2023 post-restriction viral rebound. Even more revealingly, for influenza B, the ARIMA model produced continuous, spurious fluctuations during non-epidemic periods (when true positivity was near zero), likely overreacting to variations in exogenous covariates like testing volumes or weather, while still failing to capture the true magnitude of the winter outbreak. In stark contrast, the LSTM framework's non-linear gating mechanisms successfully filtered out covariate noise during low-transmission periods while accurately tracking extreme epidemiological phase transitions. To open the computational "black box" and elucidate the predictive logic driving the LSTM networks, we conducted a SHAP analysis. The global feature importance rankings (Figures 6 E and F) revealed that the endogenous autoregressive component, historical influenza positivity rates, was the overwhelmingly dominant predictor for both subtypes. Beyond historical rates, the networks exhibited distinct, subtype-specific dependencies on meteorological and surveillance covariates. For influenza A (Figure 6 E), ambient temperature emerged as the second most impactful predictor, followed by solar radiation, the UV index, and weekly detection volumes. Conversely, for influenza B (Figure 6 F), weekly detection volumes emerged as the second most critical feature overall, followed by temperature and DOW. Crucially, these feature hierarchies confirm that the deep learning algorithm did not merely memorize spurious, weather-driven testing surges. Rather, it autonomously recognized the profound confounding impact of surveillance intensity bias. By assigning high predictive weight to weekly detection volumes and behavioral covariates, the LSTM network dynamically calibrated its epidemiological forecasts, effectively contextualizing intrinsic meteorological triggers within the rapidly shifting pandemic-era surveillance landscape.

Discussion

This study characterized the population-level relationship between meteorological conditions and influenza activity in Putian, a representative non-megacity subtropical coastal city in southeastern China, using time-series (2018–2023) of integrated surveillance and meteorological data. As an ecological investigation, it was designed to describe environment–influenza associations and to develop an interpretable short-horizon forecasting framework, rather than to infer individual-level causality. By combining DLNM-based association analysis with Bayesian-optimized, subtype-

specific LSTM forecasting, and by explicitly incorporating covariates related to testing intensity and NPI stringency, this work addresses three persistent gaps in subtropical influenza research: the isolated application of inferential and predictive tools, the limited use of principled hyperparameter optimization in LSTM-based influenza models, and the incomplete accommodation of pandemic-era non-stationarity in climate-sensitive forecasting. The findings deliver subtype-specific insights into meteorology–influenza coupling and establish a methodologically rigorous foundation for future operational surveillance systems in subtropical settings.

Unlike temperate regions, subtropical areas exhibit year-round influenza circulation characterized by considerable epidemiological heterogeneity. A Hong Kong-based study reported that local influenza A typically peaks between January and April, whereas influenza B peaks from May to July (Kim, Park, and Lee 2020). A study from Okinawa, Japan, identified analogous trends, with influenza A peaking from December to March and influenza B from April to June (Iha et al. 2016). Our six-year dataset revealed that both viruses circulated year-round in Putian, with peaks concentrated in winter-spring and additional summer activity. Influenza A predominantly peaked between December and April, whereas influenza B outbreaks frequently occurred outside this conventional window, suggesting that extreme weather fluctuations, rather than mere seasonality, act as primary outbreak triggers (Dave and Lee 2019). The infection burden was concentrated among individuals under 18 years (73.12%), with male predominance in childhood and female predominance in adulthood, consistent with heightened exposure risk in school-aged children through close-contact group settings (Jackson et al. 2013). Notably, the significant suppression of influenza positivity observed in 2020 likely reflects the impact of strict NPIs and altered population contact networks implemented during the early COVID-19 period. Collectively, these patterns motivated the subsequent quantitative characterization of subtype-specific meteorological drivers.

The meteorology-influenza relationship remains a critical question in environmental epidemiology, with ambient temperature, relative humidity, precipitation, DTR, and wind parameters known to modulate virion survival, host respiratory susceptibility, and population contact behaviors (Polozov et al. 2008; Sooryanarain and Elankumaran 2015). Consistent with the intrinsic non-linearity of these relationships, Pearson correlation coefficients Figure S2) were uniformly weak (all $r < .0.3$), reinforcing the inadequacy of linear between influenza positivity rates and meteorological variables in our cohort (Supplementary approaches. DLNM cross-basis analysis, in contrast, revealed clear non-linear exposure–response and lag–response structures that differed markedly between subtypes.

Cumulative exposure–response patterns diverged notably between subtypes. Influenza A positivity was associated with warmer conditions between 15°C and 24°C, with the cumulative relative risk peaking at 15°C (RR = 3.01, 95% CI: 1.03–8.78). Conversely, the cumulative risk for influenza B progressively declined from 8°C to 24°C, peaking sharply at the colder threshold of 8°C (RR = 26.12, 95% CI: 7.79–87.61). This distinct thermal preference contrasts with findings from Shanghai, where both subtypes showed maximal cumulative risk near 1.4°C (Zhang et al. 2020). Enhanced transmission at lower temperatures is broadly consistent with greater envelope stability in cold air (Lowen and Steel 2014), impaired mucociliary clearance (Eccles 2002), and increased indoor confinement facilitating aerosol exchange (Liao, Chang, and Liang 2005). Conversely, elevated temperatures can also inadvertently increase transmission through reduced outdoor activity and greater reliance on closed-cycle air conditioning (Lofgren et al. 2007). The subtype-specific divergence may further reflect fundamental virological and immunological differences, including the relative stability of influenza A glycoproteins under warmer, humid conditions and a structural preference of influenza B for cooler, wetter environments (Marr et al. 2019). Immunologically, high ambient temperatures can impair adaptive immune responses, potentially facilitating influenza A infection (Moriyama and Ichinohe 2019), whereas cooler settings preserve innate responses, aiding the control of influenza B (Lofgren et al. 2007).

Complex host-environment interactions may also suppress specific immune components, such as immunoglobulin M, inadvertently creating a permissive environment for influenza B transmission (Xu, Hu, and Tian 2017).

High relative humidity has been widely documented to enhance influenza risk by promoting viral transmission, as corroborated by multi-city studies in the Sichuan Basin (Zhou et al. 2022). Cellular models indicate that humidity influences respiratory droplets evaporation, altering solute concentration and influenza A virus viability (Yang, Elankumaran, and Marr 2012), while findings from Arizona indicate increased influenza risk during the rainy season (Soebiyanto, Adimi, and Kiang 2010). Consistent with these observations, our results showed that the cumulative risk for influenza B gradually escalated within a high-humidity range of 76% to 91%, peaking at 91% (RR = 1.52, 95% CI: 1.00–2.29). The cumulative risk for influenza B also progressively increased with decreasing solar radiation, peaking 10 W/m² (RR = 8.66, 95% CI: 2.14–35.00), suggesting that insufficient sunlight exposure may specifically promote influenza B transmission, a pattern absent for influenza A. Extensive literature supports that adequate solar radiation and subsequent UV exposure reduce airborne influenza viability and support host vitamin D synthesis and immune modulation (Cannell et al. 2006; Sagripanti and Lytle 2007). However, contradictory effects of UV and solar radiation on subtypes A and B reported in pediatric Vitamin D supplementation trials (Urashima et al. 2010) necessitate further mechanistic investigation into the interplay between radiant energy and specific viral lineages. Physical mechanical factors such as DTR and wind metrics also demonstrated nuanced impacts. Prior research indicates that DTR serves as a critical risk factor, with severe fluctuations (11–14°C) increasing infection risk by compromising respiratory epithelial barriers and modulating inflammatory mucus production (Park et al. 2020). Our analysis partially corroborated this, showing that narrower DTRs (1–8°C) heightened cumulative risks for both subtypes, potentially reflecting sustained susceptibility under chronically stable yet suboptimal thermal conditions (Cheng et al. 2014). Wind direction influenced only influenza B, while wind speed exerted no significant cumulative effect on either subtype, consistent with literature deeming wind speed a non-critical epidemiological variable (Ianevski et al. 2019).

Lag structures further distinguished the subtypes within the 15-day post-exposure window. Low average temperatures of 14°C were associated with a rapid 1-day lagged effect for influenza A, but induced a prolonged, bimodal lagged risk for influenza B (0–5 days and 9–10 days). Conversely, high averages of 30.5°C conferred a delayed risk for influenza B only at 4.2–7.2 days, a timeline closely mirroring Wang et al.'s reported 7.3-day lag effect for influenza B at 28°C (Wang et al. 2023). Extreme high precipitation (25 mm) induced a delayed risk for influenza B at 7.8–11 days, consistent with meta-analyses confirming that extreme rainfall events in the subtropics trigger lagged infectious disease surges (Aune, Davis, and Smith 2021). Notably, our data revealed a profound paradox regarding diurnal temperature variation. While a high DTR (13°C) exhibited significant cumulative relative risk, it produced no distinct lagged risk peaks. In stark contrast, a low DTR (3°C) triggered pronounced lagged risks for both influenza A (0–2 and 11–12.8 days) and influenza B (10.2–12.2 days). The intrinsic physiological mechanisms underlying this divergence, where broad temperature swings drive overall cumulative burden, but narrow temperature bands dictate specific temporal outbreak windows, warrant deeper investigation. Overall, these results identify the 1-to-2-week window following specific meteorological anomalies as an operationally meaningful period for anticipatory action, while emphasizing that thresholds and lag structures require local calibration. The fixed 15-day maximum lag adopted here, though consistent with prior literature, may not capture all longer-range environmental effects; systematic sensitivity analyses varying the lag window would be a valuable extension.

A central methodological feature of our framework is the deliberate division of labor between the DLNM and LSTM components. Multivariate DLNMs efficiently expose transparent, lag-resolved main-effect surfaces for each meteorological variable, but cannot accommodate high-dimensional interactions among meteorological, autoregressive, and socio-behavioral factors without parameter inflation and severe multicollinearity. The LSTM stage was therefore not intended to replace DLNM inference, but to complement it by learning joint non-linear structure across

concurrent covariates. Furthermore, extensive environmental inputs inevitably introduce severe collinearity, such as the strongly correlated solar radiation and UV index. While traditional multivariate models are highly vulnerable to such overlapping variances, the recurrent, weighted representation learned by the LSTM is comparatively tolerant of such redundancy, allowing broader covariate integration than in previous efforts (Zhu et al. 2022). Crucially, this LSTM stage explicitly incorporates weekly detection volumes, mask-wearing stringency indices, non-local population proportion, and DOW effects alongside meteorological inputs, and uses influenza positivity rates rather than absolute case counts as the modeling endpoint. Together, these design choices are intended to mitigate, rather than fully eliminate, the surveillance-related biases that can distort count-based forecasting during periods of fluctuating testing intensity.

To maximize this architectural advantage and ensure mathematical robustness, we conducted refined hyperparameter optimizations using the Hyperopt framework (Li, Zhang, and Ma 2023). By employing a Bayesian strategy based on Tree-based Parzen Estimators (TPE) (Hanifi et al. 2022), our approach efficiently navigated the complex, high-dimensional hyperparameter space. Unlike traditional heuristic or manual grid search methods, this Bayesian approach systematically optimized the learning rate, network depth, and activation functions to strike an optimal balance between model fitting capacity and the prevention of overfitting (Hanifi, Cammarono, and Zare-Behtash 2024). To our knowledge, this represents a novel application of Hyperopt-based tuning for an epidemiological LSTM forecasting network, ensuring that our algorithm's superior performance is grounded in mathematically rigorous architectural configuration rather than arbitrary parameter selection.

The optimized LSTM algorithm demonstrated strong predictive performance, with loss curves for both subtype-specific networks exhibiting favourable convergence, consistent with prior work (Du et al. 2023). LSTM consistently outperformed the covariate-matched ARIMA benchmark for both influenza A (MAE 0.009 vs 0.136) and influenza B (MAE 0.002 vs 0.049), directly addressing the methodological question of what specific value a memory-enabled deep learning architecture adds relative to a well-established linear autoregressive reference. Because the ARIMA baseline was supplied with the same meteorological and surveillance covariates as the LSTM, the observed gap is unlikely to reflect informational disparity alone; rather, ARIMA's assumption of linear additivity appears to have limited its ability to track the abrupt post-NPI rebound in 2023, whereas the LSTM accommodated this non-stationarity without architectural modification. This advantage, together with the comparative edge over previously reported ARIMA-based (Li et al. 2024) and similar LSTM-based influenza prediction models (Zhu et al. 2022), positions our framework as a substantive advance in predictive modeling for climate-sensitive diseases. External evaluation using independent Sanming CDC data yielded MAE values of 0.0152 for influenza A and 0.0054 for influenza B, providing preliminary evidence of model portability within the southeastern Chinese subtropical context, though this rests on a single external city and should not be interpreted as broad spatial generalizability. SHAP analysis (Figure 6 E–F) further enhanced LSTM interpretability, identifying historical influenza positivity rates, weekly detection volumes, temperature, and solar radiation as the most highly weighted predictive features, a pattern consistent with temperature- and radiation-mediated influenza virus inactivation mechanisms (Sagripanti and Lytle 2007), and corroborating the network's autonomous capacity to calibrate predictions against fluctuating surveillance intensities. While SHAP's underlying assumption of feature independence may not fully capture temporal interactions within the LSTM network, treating each meteorological variable's lagged sequence as a unified feature partially preserves temporal dependencies. Future work could explore temporal SHAP extensions or alternative interpretability methods designed for sequential data.

Beyond its forecasting accuracy, the DLNM-LSTM framework offers several potential contributions for subtropical public health practice. By delineating subtype-specific meteorological thresholds, characterizing their lagged temporal effects, and dynamically integrating behavioral and surveillance covariates, it enables the anticipation of high-risk transmission windows 1 to 2 weeks in advance of clinical signal emergence. Such foresight can inform pre-emptive vaccination campaigns for vulnerable populations, optimize antiviral stockpiling and healthcare resource

allocation, and intensify surveillance during predicted risk windows to accelerate case detection and containment. The underlying dual-stage architecture is also methodologically extensible to other climate-sensitive infectious diseases, including dengue and hand-foot-and-mouth disease, in which non-linear environmental coupling and lagged effects similarly dominate transmission dynamics. In a warming climate, where influenza seasonality may shift in response to evolving weather regimes, the framework's modular design permits periodic recalibration to maintain predictive validity. Importantly, however, this framework should be regarded as a methodological foundation for future operational developments rather than as a deployable early-warning system: routine use in public health practice would require prospective recalibration, integration with operational surveillance infrastructure, and additional validation beyond the scope of the present study.

This study has several limitations that should be considered when interpreting the findings. Firstly, our primary analysis relies on influenza surveillance data collected from seven sentinel hospitals in Putian, which inherently captures only a fraction of all influenza cases occurring in the broader community. Although employing positivity rates as the primary outcome substantially mitigates the surveillance bias inherent in count-based analyses, residual selection effects may persist if testing patterns shift differentially across demographic subgroups with systematically divergent positivity profiles. Our inclusion of weekly detection volumes as an explicit LSTM covariate, complemented by parallel DLNM sensitivity analyses validating the independence of meteorological effects from testing volumes, were designed to characterize and partially account for this residual bias, but cannot fully eliminate it. Consequently, positivity rates likely underestimate the true burden of community influenza infection. Although the surveillance infrastructure in Putian remained uniform and uninterrupted throughout the 2018–2023 timeline, mitigating measurement bias and temporal confounding risks, the transferability of our findings to other global subtropical regions with different socioeconomic structures, healthcare systems, or population behaviors requires cautious, region-specific calibration. Accordingly, our framework should be strictly interpreted as a high-fidelity tool designed to forecast the observable public health surveillance signal, which is the most operationally relevant target for public health agencies, rather than for estimating unobserved, absolute community disease burden. Future research should aim to integrate broader data sources, such as community-level surveillance or syndromic data streams where available, and pursue multi-centre studies across diverse global subtropical regions to disentangle core meteorological determinants that transcend geographical boundaries from those requiring local adaptation.

Secondly, the interpretation of our framework's predictive performance during the 2023 validation period requires careful epidemiological and methodological contextualization. The year of 2023 represented an anomalous, post-restriction "rebound" period characterized by rapid NPI relaxation and the release of accumulated population-level immunity debt, resulting in an atypical influenza surge that exceeded pre-pandemic peaks. The framework's high accuracy across this period should therefore be interpreted as evidence of algorithmic agility and adaptive capacity during a highly volatile transitional phase, rather than as definitive proof of long-term predictive validity under a stabilized post-2024 epidemiological regime. Methodologically, while our strict chronological OOT data partitioning prevented temporal information leakage, a critical requirement for LSTM integrity, the reliance on a single, fixed chronological split point (December 31, 2022) intrinsically limits our evaluation to one specific structural break. This fixed-split approach may not exhaustively probe the DLNM-LSTM framework's resilience against all forms of future epidemiological non-stationarity. Consequently, naive extrapolation of the reported 2023 performance metrics to future surveillance years should be avoided absent prospective recalibration. Future studies should consider employing expanding-window or rolling-origin cross-validation frameworks to provide a more continuous characterization of algorithmic robustness. Continuous integration of accumulating 2024 and 2025 data, combined with adaptive learning architectures capable of detecting regime shifts in real time, will be essential before any operational deployment of this, or similar forecasting frameworks, for routine public health surveillance.

Thirdly, beyond the surveillance coverage limitation noted above, the aggregate-level nature of the available data further constrained our ability to adjust for individual-level confounders, including personal vaccination status, detailed comorbidities, healthcare-seeking behavior, and socioeconomic status. This limitation was particularly exacerbated by the profound epidemiological disruptions during the COVID-19 pandemic, where raw numbers of confirmed cases became heavily confounded by surveillance intensity (i.e., fluctuating testing volumes) rather than solely reflecting underlying viral transmission. Our adoption of influenza positivity rates as the primary modeling endpoint and incorporation of weekly detection volumes, face-covering stringency, and non-local population proportion as dynamic covariates, rigorously mitigated these aggregate-level biases and linked our methodological design directly to the forecasting outcomes. As unequivocally demonstrated by our covariate-ablation sensitivity analyses (Table 2), failing to account for mask mandates and testing volumes leads to severe, mathematically predictable deviations in absolute forecasting accuracy. Furthermore, while daily case counts in a single city can occasionally be small, sporadic, and driven by external importations, our incorporation of non-local population proportion effectively adjusted for these localized importation risks. Consequently, although our findings characterize population-level associations between meteorological factors and influenza activity, they should not be interpreted as evidence of micro-level causal mechanisms at the individual patient level. Ultimately, the DLNM-LSTM framework's robust performance across the non-stationary transition out of NPI policies highlights the absolute necessity of integrating behavioral and virological baseline metrics into future climate-driven predictive surveillance systems.

Fourthly, although our framework demonstrates that the LSTM network substantially outperformed the covariate-adjusted ARIMA baseline, our study did not encompass a comprehensive algorithmic benchmarking against other advanced machine learning architectures. We deliberately selected ARIMA as our primary comparator because it represents the canonical, field-standard statistical reference in public health surveillance. This design choice was specifically tailored to address our primary methodological question: quantifying the predictive value gained when transitioning from a traditional linear, parametrically constrained autoregressive model to a non-linear deep learning architecture under highly volatile pandemic conditions. We fully acknowledge that modern tree-based ensemble algorithms, such as XGBoost and Random Forest, are exceptionally powerful, highly interpretable, and computationally efficient tools for epidemiological time-series forecasting. We prioritized the LSTM architecture for this specific framework because its recurrent gating mechanisms natively process sequential dependencies, allowing the model to endogenously learn optimal lag representations without the need for manual, analyst-dependent feature engineering of sliding windows. Nevertheless, the absence of tree-based baselines remains a limitation of the current evaluation. As this forecasting framework evolves toward operational deployment, systematic horizontal benchmarking against tree-based ensembles and emerging Transformer architectures will be essential to fully characterize the algorithmic performance landscape and optimize real-time scalability.

Finally, it is imperative to explicitly delineate the generalizability of our forecasting framework within geographic and epidemiological contexts. The successful external validation in Sanming provides direct empirical evidence that our LSTM architecture and the identified meteorological thresholds are robustly transferable across the subtropical southeastern Chinese context, which shares comparable monsoon climates, year-round influenza circulation, and standardized public health intervention frameworks. However, we explicitly define the boundaries of this transferability. The specific meteorological coefficients, lag structure, and predictive parameters derived in this study should not be indiscriminately extrapolated to other regions, even within subtropical latitudes, where climatic regimes (e.g., non-monsoon subtropics) or public health contexts (e.g., higher baseline influenza vaccination coverage or distinct NPI strategies) differ substantially. In such disparate settings, directly applying our pretrained model weights may yield substantial systematic biases. While the underlying dual-modeling architecture remains methodologically transferable, its parameterization requires rigorous local recalibration using

region-specific surveillance data. Acknowledging these boundaries ensures that the framework can be deployed safely and appropriately to realize targeted, climate-sensitive infectious disease forecasts.

In conclusion, this study elucidates the distinct, non-linear meteorological drivers of influenza A and B transmission in a subtropical Chinese urban setting through an integrated dual-stage DLNM-LSTM framework. By adopting influenza positivity rates as the primary outcome and integrating socio-behavioral covariates, including mask-wearing stringency indices, weekly detection volumes, and non-local population proportion indicators, our approach mitigates surveillance-related biases and accommodates pandemic-era non-stationarity. It shows lower forecast error than a covariate-matched ARIMA baseline and provides preliminary evidence of portability within southeastern subtropical China, combining the interpretability of distributed lag modeling with the flexibility of deep learning. The framework offers an interpretable, climate-informed methodological foundation for future operational surveillance developments in subtropical settings.

Data availability

The influenza surveillance data and meteorological records utilized in this study, as well as the code written for constructing the DLNM models and LSTM algorithms, are available upon reasonable request to the corresponding authors X-W. J. and M-J. Z. All data are managed in accordance with the ethical standards and data management policies of the affiliated institutions. Requests for data access will be reviewed and fulfilled in compliance with these policies and regulations. The computational codes used in this study, including those for the distributed lag non-linear models (DLNM) and the Bayesian-optimized Long Short-Term Memory (LSTM) neural network implementations, were written in R and Python. All custom scripts used for data analysis, model development, and visualization are available upon reasonable request from the corresponding authors X-W. J. and M-J. Z. The codes include data preprocessing, model training, validation procedures, and the generation of prediction results. Researchers interested in reproducing or building upon our methodology may contact the corresponding authors with specific requirements regarding the code components needed for their research purposes.

Acknowledgements

We acknowledge the invaluable contribution of the Putian CDC and its affiliated sentinel hospitals to this study. We express our deep respect and gratitude to the dedication of all participating healthcare workers, disease prevention personnel, clinical medical staff, and laboratory technicians involved in case screening, data recording, sample collection, and laboratory testing for their significant efforts and contributions to this study.

Additional information

Author Contributions

Conceptualisation: X-W. J., L-M. L., X. H., S. X., and L. X; Data curation: M-J. Z., J-L. T., Z-S. W., Y-X. L., J-L. C., J-J. H., J. Q., and H-Y. P; Formal analysis: L. X., M-J. Z., J-L. T., Z-S. W., Y-X. L., Z-Q. X., S. X., and X. H; Funding acquisition: L-M. L. and X-W. J; Investigation: M-J. Z., Y-X. L., J. Q., J-J. H., J-L. C., H-Y. P., and Z-F. R; Methodology: J-L. T., Z-S. W., L. X., X-W. J., X. H., S. X., and M-J. Z; Software: J. Q., H-Y. P., J-L. T., Z-S. W., and Z-F. R; Project administration: X-W. J., L-M. L., and X. H; Resource: X-W. J, L-M. L., S. X., and X. H; Supervision: X-W. J., L-M. L., M-J. Z., and Y-X. L.; Validation: L. X., Y-X. L., J-L. T., Z-S. W., Z-Q. X., X. H., and M-J. Z; Visualisation: L. X., J-L. T., Z-S. W.,

J-L. C., J. Q., J-J. H., and Z-F. R; Writing-original draft: L. X., and J-L. T; Writing-review & editing: L. X., X-W. J., L-M. L., S. X., and X. H. All authors have reviewed and approved the final version of the manuscript.

Funding

The study was financially supported by a grant from Guangdong Provincial Science and Technology Program (grant number: 2022B1111020003) awarded to Xi-Wen Jiang. Our study was carried out with complete autonomy, with the assurance that the funding bodies had no influence at any stage of the research process.

Funding

Funder	Grant reference number	Author
Guangdong Provincial Department of Science and Technology (GDSTC)	2022B1111020003	Xiwen Jiang

Author ORCID iDs

Long Xie: https://orcid.org/0000-0002-2138-888X
Xi-Wen Jiang: https://orcid.org/0009-0002-7977-4510

Additional files

Supplemental Information. Table S1: Descriptive Statistics of Meteorological Factors in Putian City (2018-2023). Table S2: QAIC Values for DLNM Models Across Degrees of Freedom Combinations. Figure S1: Detailed Structure of the Long Short-Term Memory (LSTM) Cell.Figure S2: Correlation Heatmap and Scatterplot Matrix Illustrating Pairwise Associations and Potential Collinearity among Influenza Indicators and Meteorological Variables.Figure S3. Sensitivity Analysis of the Cumulative Risk of Meteorological Factors on Influenza A and B Positivity Rates.Figure S4. Sensitivity Analysis of the Lagged Temporal Effects of Extreme Meteorological Conditions on Influenza A and B Positivity Rates.Figure S5. The Actual Values of Influenza A and B Compared to the Predicted Values from the Covariate-Adjusted ARIMA Model.

References

1. **Absar N,** Uddin N, Khandaker MU, Ullah H (2022) The efficacy of deep learning based LSTM model in forecasting the outbreak of contagious diseases. *Infect Dis Model* **7**:170-83 https://doi.org/10.1016/j.idm.2021.12.005 | PubMed
2. **Aune KT,** Davis MF, Smith GS (2021) Extreme Precipitation Events and Infectious Disease Risk: A Scoping Review and Framework for Infectious Respiratory Viruses. *Int J Environ Res Public Health* **19** https://doi.org/10.3390/ijerph19010165 | PubMed
3. **Bloom-Feshbach K,** Alonso WJ, Charu V, Tamerius J, Simonsen L, Miller MA, Viboud C (2013) Latitudinal variations in seasonal activity of influenza and respiratory syncytial virus (RSV): a global comparative review. *PLoS One* **8**:e54445 https://doi.org/10.1371/journal.pone.0054445 | PubMed
4. **Cannell JJ,** Vieth R, Umhau JC, Holick MF, Grant WB, Madronich S, Garland CF, Giovannucci E (2006) Epidemic influenza and vitamin D. *Epidemiol Infect* **134**:1129-40 https://doi.org/10.1017/s0950268806007175 | PubMed
5. **Cheng J,** Xu Z, Zhu R, Wang X, Jin L, Song J, Su H (2014) Impact of diurnal temperature range on human health: a systematic review. *Int J Biometeorol* **58**:2011-24 https://doi.org/10.1007/s00484-014-0797-5 | PubMed
6. **Chicco D,** Warrens MJ, Jurman G (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput Sci* **7**:e623 https://doi.org/10.7717/peerj-cs.623 | PubMed

7. **Dave K,** Lee PC (2019) Global Geographical and Temporal Patterns of Seasonal Influenza and Associated Climatic Factors. *Epidemiol Rev* **41**:51-68 https://doi.org/10.1093/epirev/mxz008 | PubMed
8. **Deyle ER,** Maher MC, Hernandez RD, Basu S, Sugihara G (2016) Global environmental drivers of influenza. *Proc Natl Acad Sci U S A* **113**:13081-86 https://doi.org/10.1073/pnas.1607747113 | PubMed
9. **Du X,** Si L, Li P, Yun Z (2023) A method for detecting the quality of cotton seeds based on an improved ResNet50 model. *PLoS One* **18**:e0273057 https://doi.org/10.1371/journal.pone.0273057 | PubMed
10. **Eccles R** (2002) An explanation for the seasonality of acute upper respiratory tract viral infections. *Acta Otolaryngol* **122**:183-91 https://doi.org/10.1080/00016480252814207 | PubMed
11. **Gasparrini A** (2011) Distributed Lag Linear and Non-Linear Models in R: The Package dlnm. *J Stat Softw* **43**:1-20 PubMed
12. **Gasparrini A,** Armstrong B, Kenward MG (2010) Distributed lag non-linear models. *Stat Med* **29**:2224-34 https://doi.org/10.1002/sim.3940 | PubMed
13. **Groves HE,** Piché-Renaud PP, Peci A, Farrar DS, Buckrell S, Bancej C, Sevenhuysen C, Campigotto A, Gubbay JB, Morris SK (2021) The impact of the COVID-19 pandemic on influenza, respiratory syncytial virus, and other seasonal respiratory virus circulation in Canada: A population-based study. *Lancet Reg Health Am* **1**:100015 https://doi.org/10.1016/j.lana.2021.100015 | PubMed
14. **Guo Q,** Dong Z, Zeng W, Ma W, Zhao D, Sun X, Gong S, Xiao J, Li T, Hu W (2019) The effects of meteorological factors on influenza among children in Guangzhou, China. *Influenza Other Respir Viruses* **13**:166-75 https://doi.org/10.1111/irv.12617 | PubMed
15. **Hanifi Shahram,** Cammarono Andrea, Zare-Behtash Hossein (2024) Advanced hyperparameter optimization of deep learning models for wind power prediction. *Renewable Energy* **221**:119700 https://doi.org/10.1016/j.renene.2023.119700
16. **Hanifi Shahram,** Lotfian Saeid, Zare-Behtash Hossein, Cammarano Andrea (2022) Offshore Wind Power Forecasting—A New Hyperparameter Optimisation Algorithm for Deep Learning Models. *Energies* **15**:6919 https://doi.org/10.3390/en15196919
17. **Ianevski A,** Zusinaite E, Shtaida N, Kallio-Kokko H, Valkonen M, Kantele A, Telling K, Lutsar I, Letjuka P, Metelitsa N, *et al.* (2019) Low Temperature and Low UV Indexes Correlated with Peaks of Influenza Virus Activity in Northern Europe during 2010L2018. *Viruses* **11** https://doi.org/10.3390/v11030207 | PubMed
18. **Iha Y,** Kinjo T, Parrott G, Higa F, Mori H, Fujita J (2016) Comparative epidemiology of influenza A and B viral infection in a subtropical region: a 7-year surveillance in Okinawa, Japan. *BMC Infect Dis* **16**:650 https://doi.org/10.1186/s12879-016-1978-0 | PubMed
19. **Iuliano AD,** Roguski KM, Chang HH, Muscatello DJ, Palekar R, Tempia S, Cohen C, Gran JM, Schanzer D, Cowling BJ, *et al.* (2018) Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *Lancet* **391**:1285-300 https://doi.org/10.1016/s0140-6736(17)33293-2 | PubMed
20. **Jackson C,** Vynnycky E, Hawker J, Olowokure B, Mangtani P (2013) School closures and influenza: systematic review of epidemiological studies. *BMJ Open* **3** https://doi.org/10.1136/bmjopen-2012-002149 | PubMed
21. **Javanian M,** Barary M, Ghebrehewet S, Koppolu V, Vasigala V, Ebrahimpour S (2021) A brief review of influenza virus infection. *J Med Virol* **93**:4638-46 https://doi.org/10.1002/jmv.26990 | PubMed
22. **Kim BI,** Park O, Lee S (2020) Comparison of influenza surveillance data from the Republic of Korea, selected northern hemisphere countries and LHong Kong Special Administrative Region SAR (China) from 2012 to 2017. *Western Pac Surveill Response J* **11**:1-9 https://doi.org/10.5365/wpsar.2019.10.2.015 | PubMed
23. **Li G,** Li Y, Han G, Jiang C, Geng M, Guo N, Wu W, Liu S, Xing Z, Han X, *et al.* (2024) Forecasting and analyzing influenza activity in Hebei Province, China, using a CNN-LSTM hybrid model. *BMC Public Health* **24**:2171 https://doi.org/10.1186/s12889-024-19590-8 | PubMed

24. **Li Q,** Zhang Z, Ma Z (2023) Raman spectral pattern recognition of breast cancer: A machine learning strategy based on feature fusion and adaptive hyperparameter optimization. *Heliyon* **9**:e18148 https://doi.org/10.1016/j.heliyon.2023.e18148 | PubMed

25. **Li Y,** Reeves RM, Wang X, Bassat Q, Brooks WA, Cohen C, Moore DP, Nunes M, Rath B, Campbell H, *et al.* (2019) Global patterns in monthly activity of influenza virus, respiratory syncytial virus, parainfluenza virus, and metapneumovirus: a systematic analysis. *Lancet Glob Health* **7**:e1031-e45 https://doi.org/10.1016/s2214-109x(19)30264-5 | PubMed

26. **Liao CM,** Chang CF, Liang HM (2005) A probabilistic transmission dynamic model to assess indoor airborne infection risks. *Risk Anal* **25**:1097-107 https://doi.org/10.1111/j.1539-6924.2005.00663.x | PubMed

27. **Liu J,** Chen E, Zhang Q, Shi P, Gao Y, Chen Y, Liu W, Qin Y, Shen Y, Shi C (2020) The correlation between atmospheric visibility and influenza in Wuxi city, China. *Medicine* **99**:e21469 https://doi.org/10.1097/md.0000000000021469 | PubMed

28. **Liu X,** Liu C, Huang R, Zhu H, Liu Q, Mitra S, Wang Y (2021) Long short-term memory recurrent neural network for pharmacokinetic-pharmacodynamic modeling. *Int J Clin Pharmacol Ther* **59**:138-46 https://doi.org/10.5414/cp203800 | PubMed

29. **Lofgren E,** Fefferman NH, Naumov YN, Gorski J, Naumova EN (2007) Influenza seasonality: underlying causes and modeling theories. *J Virol* **81**:5429-36 https://doi.org/10.1128/jvi.01680-06 | PubMed

30. **Lowen AC,** Steel J (2014) Roles of humidity and temperature in shaping influenza seasonality. *J Virol* **88**:7692-5 https://doi.org/10.1128/jvi.03544-13 | PubMed

31. **Marr LC,** Tang JW, Van Mullekom J, Lakdawala SS (2019) Mechanistic insights into the effect of humidity on airborne influenza virus survival, transmission and incidence. *J R Soc Interface* **16**:20180298 https://doi.org/10.1098/rsif.2018.0298 | PubMed

32. **Moriyama M,** Hugentobler WJ, Iwasaki A (2020) Seasonality of Respiratory Viral Infections. *Annu Rev Virol* **7**:83-101 https://doi.org/10.1146/annurev-virology-012420-022445 | PubMed

33. **Moriyama M,** Ichinohe T (2019) High ambient temperature dampens adaptive immune responses to influenza A virus infection. *Proc Natl Acad Sci U S A* **116**:3118-25 https://doi.org/10.1073/pnas.1815029116 | PubMed

34. **Ng H,** Li Y, Zhang T, Lu Y, Wong C, Ni J, Zhao Q (2022) Association between multiple meteorological variables and seasonal influenza A and B virus transmission in Macau. *Heliyon* **8**:e11820 https://doi.org/10.1016/j.heliyon.2022.e11820 | PubMed

35. **Park JE,** Son WS, Ryu Y, Choi SB, Kwon O, Ahn I (2020) Effects of temperature, humidity, and diurnal temperature range on influenza incidence in a temperate region. *Influenza Other Respir Viruses* **14**:11-18 https://doi.org/10.1111/irv.12682 | PubMed

36. **Peci A,** Winter AL, Li Y, Gnaneshan S, Liu J, Mubareka S, Gubbay JB (2019) Effects of Absolute Humidity, Relative Humidity, Temperature, and Wind Speed on Influenza Activity in Toronto, Ontario, Canada. *Appl Environ Microbiol* **85** https://doi.org/10.1128/aem.02426-18 | PubMed

37. **Polozov IV,** Bezrukov L, Gawrisch K, Zimmerberg J (2008) Progressive ordering with decreasing temperature of the phospholipids of influenza virus. *Nat Chem Biol* **4**:248-55 https://doi.org/10.1038/nchembio.77 | PubMed

38. **Ryu S,** Cowling BJ (2021) Human Influenza Epidemiology. *Cold Spring Harb Perspect Med* **11** https://doi.org/10.1101/cshperspect.a038356 | PubMed

39. **Sagripanti JL,** Lytle CD (2007) Inactivation of influenza virus by solar radiation. *Photochem Photobiol* **83**:1278-82 https://doi.org/10.1111/j.1751-1097.2007.00177.x | PubMed

40. **Shoji M,** Katayama K, Sano K (2011) Absolute humidity as a deterministic factor affecting seasonal influenza epidemics in Japan. *Tohoku J Exp Med* **224**:251-6 https://doi.org/10.1620/tjem.224.251 | PubMed

41. **Si Xiaohan,** Mengersen Kerrie, Ye Chuchu, Hu Wenbiao (2024) Interactive effect of air pollutant and meteorological factors on seasonal influenza transmission, Shanghai, China. *Atmospheric Environment* **318**:120208 https://doi.org/10.1016/j.atmosenv.2023.120208

42. **Soebiyanto RP,** Adimi F, Kiang RK (2010) Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters. *PLoS One* **5**:e9450 https://doi.org/10.1371/journal.pone.0009450 | PubMed

43. **Soebiyanto RP,** Gross D, Jorgensen P, Buda S, Bromberg M, Kaufman Z, Prosenc K, Socan M, Vega Alonso T, Widdowson MA, *et al.* (2015) Associations between Meteorological Parameters and Influenza Activity in Berlin (Germany), Ljubljana (Slovenia), Castile and León (Spain) and Israeli Districts. *PLoS One* **10**:e0134701 https://doi.org/10.1371/journal.pone.0134701 | PubMed

44. **Sooryanarain H,** Elankumaran S (2015) Environmental role in influenza virus outbreaks. *Annu Rev Anim Biosci* **3**:347-73 https://doi.org/10.1146/annurev-animal-022114-111017 | PubMed

45. **Tamerius JD,** Shaman J, Alonso WJ, Bloom-Feshbach K, Uejio CK, Comrie A, Viboud C (2013) Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. *PLoS Pathog* **9**:e1003194 https://doi.org/10.1371/journal.ppat.1003194 | PubMed

46. **Urashima M,** Segawa T, Okazaki M, Kurihara M, Wada Y, Ida H (2010) Randomized trial of vitamin D supplementation to prevent seasonal influenza A in schoolchildren. *Am J Clin Nutr* **91**:1255-60 https://doi.org/10.3945/ajcn.2009.29094 | PubMed

47. **Uyeki TM,** Hui DS, Zambon M, Wentworth DE, Monto AS (2022) Influenza. *Lancet* **400**:693-706 https://doi.org/10.1016/s0140-6736(22)00982-5 | PubMed

48. **Venna SR,** Tavanaei Q, Gottumukkala RN, Raghavan VV, Maida AS, Nichols S (2018) A Novel Data-driven Model for Real-Time Influenza Forecasting. *IEEE Access* https://doi.org/10.1109/access.2018.2888585

49. **Wang C,** Li Y, Feng W, Liu K, Zhang S, Hu F, Jiao S, Lao X, Ni H, Xu G (2017) Epidemiological Features and Forecast Model Analysis for the Morbidity of Influenza in Ningbo, China, 2006-2014. *Int J Environ Res Public Health* **14** https://doi.org/10.3390/ijerph14060559 | PubMed

50. **Wang D,** Lei H, Wang D, Shu Y, Xiao S (2023) Association between Temperature and Influenza Activity across Different Regions of China during 2010-2017. *Viruses* **15** https://doi.org/10.3390/v15030594 | PubMed

51. **Xiao J,** Gao M, Huang M, Zhang W, Du Z, Liu T, Meng X, Ma W, Lin S (2022) How do El Niño Southern Oscillation (ENSO) and local meteorological factors affect the incidence of seasonal influenza in New York state. *Hyg Environ Health Adv* **4** https://doi.org/10.1016/j.heha.2022.100040 | PubMed

52. **Xu DL,** Hu XK, Tian YF (2017) Effect of temperature and food restriction on immune function in striped hamsters (Cricetulus barabensis). *J Exp Biol* **220**:2187-95 https://doi.org/10.1242/jeb.153601 | PubMed

53. **Yang W,** Elankumaran S, Marr LC (2012) 'Relationship between humidity and influenza A viability in droplets and implications for influenza's seasonality'. *PLoS One* **7**:e46789 https://doi.org/10.1371/journal.pone.0046789 | PubMed

54. **Zhang R,** Peng Z, Meng Y, Song H, Wang S, Bi P, Li D, Zhao X, Yao X, Li Y (2022) Temperature and influenza transmission: Risk assessment and attributable burden estimation among 30 cities in China. *Environ Res* **215**:114343 https://doi.org/10.1016/j.envres.2022.114343 | PubMed

55. **Zhang Y,** Ye C, Yu J, Zhu W, Wang Y, Li Z, Xu Z, Cheng J, Wang N, Hao L, *et al.* (2020) The complex associations of climate variability with seasonal influenza A and B virus transmission in subtropical Shanghai, China. *Sci Total Environ* **701**:134607 https://doi.org/10.1016/j.scitotenv.2019.134607 | PubMed

56. **Zhao Z,** Zhai M, Li G, Gao X, Song W, Wang X, Ren H, Cui Y, Qiao Y, Ren J, *et al.* (2023) Study on the prediction effect of a combined model of SARIMA and LSTM based on SSA for influenza in Shanxi Province, China. *BMC Infect Dis* **23**:71 https://doi.org/10.1186/s12879-023-08025-1 | PubMed

57. **Zhou L,** Yang H, Pan W, Xu J, Feng Y, Zhang W, Shao Z, Li T, Li S, Huang T, *et al.* (2022) Association between meteorological factors and the epidemics of influenza (sub)types in a subtropical basin of Southwest China. *Epidemics* **41**:100650 https://doi.org/10.1016/j.epidem.2022.100650 | PubMed

58. **Zhu H,** Chen S, Lu W, Chen K, Feng Y, Xie Z, Zhang Z, Li L, Ou J, Chen G (2022) Study on the influence of meteorological factors on influenza in different regions and predictions based on an LSTM algorithm. *BMC Public Health* **22**:2335 https://doi.org/10.1186/s12889-022-14299-y | PubMed

Peer reviews

Reviewer #2 (Public review):

Summary:

The study aimed to assess the associations between meteorological drivers and influenza is important although not new. The authors used 6 years of surveillance data and deep learning models, combining distributed lag non-linear models (DLNM) with Bayesian-optimized LSTM neural networks for predictive modeling. The key interest in this area is to explore the subtropical locations, where influenza is less common and circulates year-round. The authors further claimed that such an association could be able to provide an early warning in the community.

Strengths:

Study design based on a prospective cohort to analyse the data for retrospective outcomes.

[Editors' note: The Reviewing Editor has assessed the revised article without further input from the original reviewers. The authors have made considerable efforts to address the methodological and reporting concerns raised by the initial peer review, resulting in a substantially revised manuscript.]

https://doi.org/10.7554/eLife.107767.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***eLife Assessment***
>
> *This study is a valuable contribution to the evidence base. However, the evidence provided is incomplete as the study results only partially support the study conclusions. Addressing the methodological and reporting issues raised by the peer reviewers and properly aligning the claim made for providing a tool for early warning with the study analysis/results would improve the study quality and usefulness of its findings.*

We are deeply encouraged by the editors' recognition of this study as a valuable contribution to the evidence base. We fully concur with the eLife assessment that the manuscript, in its original form, required substantive methodological recalibration and rhetorical refinement to ensure the claims were strictly supported by the data. We accept these profound critiques unreservedly and have undertaken a comprehensive, ground-up revision of this study.

This revision was guided by three overarching principles: (1) systematic decoupling of epidemiological confounders, (2) rigorous elimination of COVID-19-era surveillance bias, and (3) precise alignment of our claims with the actual scope of our predictive framework. Most importantly, we have fundamentally recalibrated our core claim: we have entirely removed all assertions of providing a "ready-to-use early warning tool." Instead, we accurately reposition our contribution as developing a "robust, climate-informed predictive surveillance

framework" that establishes an evidence-based foundation for understanding non-stationary climate-influenza dynamics in a subtropical urban setting.

To ensure our analytical results definitively support this recalibrated conclusion, we executed a comprehensive remodeling of our entire dataset, rebuilding both the DLNM and LSTM networks. The major structural upgrades include:

(1) Shift to Positivity Rate to Eliminate Testing Bias: To directly address the reviewers' incisive concern regarding testing volume bias (i.e., raw case counts were artificially inflated or suppressed by massive fluctuations in PCR testing during the COVID-19 pandemic), we have fundamentally replaced "influenza positive case counts" with the "influenza test positivity rates" as our primary outcome metric. This relative metric mathematically standardizes the denominator, elegantly isolating intrinsic viral transmissibility from the severe artificial fluctuations of healthcare-seeking behaviors and diagnostic intensity.

(2) Integration of New Epidemiological Covariates: We significantly enhanced our control for critical confounders (as suggested by the reviewers) by incorporating new, highly relevant covariates into our models. These include:

- Weekly detection volumes: To capture residual testing capacity fluctuations.

- Mask-wearing stringency indices (OWID): To computationally decouple the artificial suppression of cases caused by strict non-pharmaceutical interventions (NPIs).

- Proportion of the non-local/transient population: To explicitly account for external viral importation risks.

- Day of the Week (DOW, weekday vs. weekend): To reflect societal mobility and administrative patterns.

Simultaneously, we removed redundant variables (e.g., raw COVID-19 case numbers) to minimize noise.

(3) Comprehensive Re-optimization, Retraining, and Sensitivity Analyses: Following this massive feature engineering, we re-executed the Bayesian hyperparameter tuning process and completely retrained the LSTM networks. We also conducted rigorous sensitivity analyses (detailed in the new Table 2 and Supplementary files) by systematically ablating variables like testing volume and mask-wearing indices, definitively proving the necessity of these covariates in our framework. All corresponding codes, main figures (Figures 1– 6), supplementary figures, and statistical tables have been entirely updated.

(4) Streamlining the Manuscript: We acknowledge the reviewers' observation that the original methodology was overly meandering. In this revision, we have ruthlessly streamlined the text. We condensed routine laboratory protocols, reorganized the Methods to logically present the "Study area" prior to the "Study design," and thoroughly rewrote the Discussion to weave our limitations directly into the interpretation of the results. This ensures conciseness, logical flow, and a sharp focus on the central narrative, tailored for the broad and rigorous readership of eLife.

We believe these deep methodological revisions have profoundly elevated the scientific rigor, interpretability, and integrity of our study. Below, we provide detailed, point-by-point responses outlining how each specific comment was addressed.

> ***Reviewer #1 (Public review):***
>
> *A major concern is that the model is trained in the midst of the COVID-19 pandemic and its associated restrictions and validated on 2023 data. The situation before, during, and after COVID is fluid, and one may not be representative of the other. The situation in*

> *2023 may also not have been normal and reflective of 2024 onward, both in terms of the amount of testing (and positives) and measures taken to prevent the spread of these types of infections. A further worry is that the retrospective prospective split occurred in October 2020, right in the first year of COVID, so it will be impossible to compare both cohorts to assess whether grouping them is sensible.*

We deeply appreciate this astute epidemiological critique. You have precisely identified the most formidable methodological challenge in pandemic-era time series modeling: the profound non-stationarity of influenza dynamics spanning the 2018–2023 timeline, driven by intensive Non-Pharmaceutical Interventions (NPIs), pandemic-related behavioral shifts, volatile testing volumes, and subsequent immunity debt. We fully agree that 2023 represented an atypical post-restriction "rebound" year and is unlikely to be straightforwardly representative of a stabilized post-2024 epidemiological steady state.

First, we wish to clarify a minor but crucial methodological detail regarding the October 2020 split. You understandably raised concerns about comparing "both cohorts." We must emphasize that there is no change in the cohort or the data collection methodology. The surveillance system, sentinel hospitals, and diagnostic protocols (managed by Putian CDC) remained identical and uninterrupted from 2018 to 2023. Although the original manuscript labelled the January 2018 – October 2020 and October 2020 – December 2023 phases as "retrospective" and "prospective" cohorts respectively, this distinction reflected only the administrative timing of ethical approval. It does not indicate any changes in sentinel hospital locations, ILI case definitions, specimen collection procedures, RT-PCR diagnostic protocols, or laboratory quality control. The patient population and clinical criteria are completely homogeneous. We have revised the "Ethics statement" subsection to eliminate this semantic confusion.

Upon receiving this insightful feedback, our team conducted extensive mathematical evaluation on how best to address this timeline heterogeneity. We initially considered formal stratified temporal analyses (e.g., splitting models into strict Pre-COVID 2018-2019, COVID-disruption 2020-2022, and Post-COVID 2023 periods) and implementing a rolling-window validation scheme. However, after careful evaluation, we concluded that strict time-slicing of this particular dataset would introduce its own substantial methodological problems more severe than the heterogeneity it sought to resolve:

(1) Statistical power constraints on for stratified Non-linear Lag analysis: The DLNM architecture requires continuous, robust longitudinal data to stably estimate two-dimensional exposure–lag–response surfaces with natural cubic splines, which typically demand approximately 16–25 effective degrees of freedom across the joint exposure × lag space. Stratifying our six-year dataset into the three intervals list above would yield:

- Pre-COVID stratum (2018–2019): ~730 daily observations and only ~200 influenza B positive events, insufficient to stably identify cross-basis surfaces, with confidence intervals expected to widen to non-informative ranges.

- COVID-disruption stratum (2020–2022): influenza circulation was substantially suppressed (though, importantly, not eliminated, a point we return to below), reducing signal density and risking that estimated surfaces reflect suppression dynamics rather than climate–transmission relationships.

- Post-restriction stratum (2023): a single year cannot independently support a DLNM with meaningful lag structure given the typical 0–14-day lag window we examine.

(2) The "training-domain contamination" problem in rolling-window validation: We carefully examined whether expanding-window rolling validation (e.g., train 2018–2020, validate 2021; train 2018–2021, validate 2022; …) would resolve the non-stationarity concern. We concluded

it would not, for a subtle but consequential reason: each such window straddles the abrupt NPI transitions, meaning that within any single training window, the model is exposed to a nonstationary mixture of regimes without any explicit signal indicating which regime each observation belongs to. Implementing a rolling window in this specific context forces the algorithm to repeatedly train on fragmented, incomplete phases of this epidemiological cycle. The model is therefore likely to learn a confounded representation in which the climate signal is partially absorbed into the implicit regime-shift signal, a pathology that is in some respects more difficult to diagnose than that of unified modelling. By employing a single, continuous 88% training block (2018–2022), we structurally guarantee that the LSTM's memory cell is exposed to the complete sequence of regime shifts, from pre-pandemic natural baseline, through extreme suppression, and ending right at the brink of the NPI relaxation. Reserving the entirely unseen 2023 "rebound year" as a chronological hold-out thereby serves as the ultimate extreme stress-test of the network's capacity to dynamically synthesize NPI relaxation signals and climate variables to forecast a historically unprecedented surge.

(3) Architectural mismatch between time-slicing and the LSTM's design principle: A core motivation for adopting an LSTM rather than period-stratified statistical models was precisely to leverage its long-term dependency memory mechanism, which is designed to allow a unified architecture to learn how predictive relationships are modulated by time-varying contextual conditions. Presegmenting the data into NPI-defined strata forecloses this principal architectural advantage and would, in effect, reduce the analysis to a series of disconnected period-specific models, a design for which the LSTM's complexity provides no benefit over simpler approaches.

(4) Tension with the surveillance-bias correction: As we discuss in our response to Public review para 2 below, a separate and equally important critique from you concerns surveillance bias from temporally varying testing intensity. Stratified analysis would compound this problem, because each stratum carries a distinct testing-intensity profile (notably the surveillance surge during 2020–2022), and stratum-specific models cannot leverage cross-period testing-volume normalization. A unified model with explicit testing-volume covariates is protective against bias than period-stratified alternatives.

Our Methodological Solution: Covariate-Driven Adaptive Learning

Instead of artificially fracturing the timeline, we substantially restructured the analytical framework to teach the model how to contextualize the pandemic-era disruption while preserving longitudinal continuity:

- First, we shifted the predictive target to Influenza Positivity Rates (laboratory confirmed cases ÷ ILI specimens tested): The positivity rate is the WHO recommended sentinel surveillance metric and intrinsically corrects for the substantial fluctuations in testing volumes, healthcare-seeking behaviour, and surveillance intensity that characterized the pre-pandemic, pandemic, and post-restriction periods, rendering the outcome metric far more comparable across the timeline.

- Second, we integrated explicit Epidemiological Context Covariates: We structurally upgraded the LSTM network by feeding it vital time-varying covariates alongside meteorological data. Specifically, we incorporated the Our World in Data (OWID) mask-wearing stringency indices, weekly detection volumes, day of the week (DOW, distinguish between weekdays and weekends to capture administrative reporting patterns) and the proportion of non-local residents among tested patients (to account for population mobility and viral importation risks).

- Third, we performed a covariate-ablation sensitivity analysis: To prove the model actively utilizes these contextual signals, we performed targeted ablation studies (new Table 2). Removing mask-wearing stringency indices increased the 2023 influenza A forecast MAE to

0.012 and the influenza B forecast MAE to 0.003 (compared with MAEs of 0.009 and 0.002, respectively, obtained when the covariate was retained), indicating a decline in the network's predictive accuracy. Removing the weekly testing volume had an even more catastrophic impact on Influenza B, surging the MAE by 100% and SMAPE by 199.6%. This provides empirical evidence that the unified network does not passively average across regimes, but dynamically leverages NPI and surveillance signals to adapt to non-stationarity. We have added a paragraph to the revised Discussion interpreting these findings as suggestive (though not definitive) evidence that the unified-with covariates architecture is contextualizing rather than averaging across regimes.

By explicitly providing the LSTM with these explicit contextual parameters, the network autonomously learned the "regime shifts." It learned that when the mask-wearing index is high, transmission is dampened despite favorable meteorological conditions.

Re-interpreting the 2023 Validation

Guided by your critique, we fully agree that 2023 was an anomalous "rebound" year and not a steady-state reflection of a post-2024 "normal." We have completely abandoned the framing that our model represents a steady-state tool for the future. Instead, we now explicitly frame the 2023 validation as an "extreme epidemiological stress test." The revised framing rests on three explicit acknowledgements:

- The 2023 validation tests forecasting performance during a transitional, post-restriction rebound year, and is best understood as an extreme stress test of the framework's adaptive capacity, not as evidence of long-term predictive validity under stabilized future conditions.

- Performance metrics observed in 2023 should not be naively extrapolated to 2024 and beyond.

- Continuous prospective recalibration as 2024–2025 data accumulate, ideally combined with adaptive learning approaches capable of detecting regime shifts in real time, will be essential before any operational deployment.

The fact that our updated LSTM network accurately forecasted the explosive, atypical 2023 viral rebound, despite being trained heavily on the suppressed 2020-2022 data, demonstrates its robust capacity to synthesize climate variables and NPI relaxation signals.

We have substantially rewritten the Discussion section to honestly acknowledge this limitation and contextualize the 2023 results, stating that continuous recalibration will be essential for future forecasting.

We invite you to review the recently included sentences in the relevant subsections as specified above.

"A major methodological strength of this study lies in its robust, uninterrupted longitudinal data collection framework spanning January 1, 2018, to December 31, 2023. While the analytical timeline encompasses a "retrospective" phase (January 1, 2018 – October 13, 2020) prior to formal ethical approval, and a "prospective" phase thereafter, we emphasize that this distinction represents a purely administrative demarcation regarding the timing of ethical approval. It does not reflect any shift in demographic cohorts, sentinel hospital locations, or data collection methodologies. Importantly, the historical data (2018– 2020) were not subjected to the recall biases or misclassification risks typical of traditional retrospective chart reviews. Rather, they were systematically extracted from a continuously operating, highly standardized public health sentinel surveillance network. From the inception of data collection through the end of 2023, the local CDC maintained absolute uniformity in clinical influenza-like illness (ILI) definitions, nasopharyngeal swabbing procedures, and real-time reverse transcription polymerase chain reaction (RT-PCR) diagnostic assays. Consequently,

the pre-2020 data possess the high-fidelity characteristics of a strict prospective cohort, ensuring unparalleled longitudinal consistency and mitigating temporal measurement bias across the entire pre-pandemic, pandemic, and postrestriction timeline." (Methods, page 8-9)

Secondly, the interpretation of our framework's predictive performance during the 2023 validation period requires careful epidemiological and methodological contextualization. The year of 2023 represented an anomalous, post-restriction "rebound" period characterized by rapid NPI relaxation and the release of accumulated population-level immunity debt, resulting in an atypical influenza surge that exceeded pre-pandemic peaks. The framework's high accuracy across this period should therefore be interpreted as evidence of algorithmic agility and adaptive capacity during a highly volatile transitional phase, rather than as definitive proof of long-term predictive validity under a stabilized post-2024 epidemiological regime. Methodologically, while our strict chronological OOT data partitioning prevented temporal information leakage, a critical requirement for LSTM integrity, the reliance on a single, fixed chronological split point (December 31, 2022) intrinsically limits our evaluation to one specific structural break. This fixed-split approach may not exhaustively probe the DLNM-LSTM framework's resilience against all forms of future epidemiological non-stationarity. Consequently, naive extrapolation of the reported 2023 performance metrics to future surveillance years should be avoided absent prospective recalibration. Future studies should consider employing expanding-window or rolling-origin cross-validation frameworks to provide a more continuous characterization of algorithmic robustness. Continuous integration of accumulating 2024 and 2025 data, combined with adaptive learning architectures capable of detecting regime shifts in real time, will be essential before any operational deployment of this, or similar forecasting frameworks, for routine public health surveillance." (Discussion, page 44-45)

We believe this integrated, covariate-based approach maintains the mathematical integrity of the time-series analysis while fully addressing your valid concerns regarding epidemiological non-stationarity.

> *The outcome of interest is the number of confirmed influenza cases. This is not only a function of weather, but also of the amount of testing. The amount of testing is also a function of historical patterns. This poses the real risk that the model confirms historical opinions through increased testing in those higher-risk periods. Of course, the models could also be run to see how meteorological factors affect testing and the percentage of positive tests. The results only deal with the number of positive (only the overall number of tests is noted briefly), which means there is no way to assess how reasonable and/or variable these other measures are. This is especially concerning as there was massive testing for respiratory viruses during COVID in many places, possibly including China.*

We are exceptionally grateful for this incisive methodological observation. You have accurately identified a fundamental validity threat, "surveillance intensity bias", that inherently constrains much of the existing climate–infectious disease literature. We fully concur with your assessment that raw case counts are jointly determined by underlying viral transmission and dynamic testing intensity. Furthermore, we recognize your highly valid concern regarding the risk of "circular reasoning," whereby meteorological factors might simply trigger higher clinical suspicion and testing rates rather than genuine transmission events—a bias severely exacerbated by the massive respiratory testing surges during the COVID-19 pandemic. To systematically dismantle this threat and address your specific recommendations, we executed a ground-up restructuring of our analytical framework, implementing four complementary strategies:

(1) Primary outcome redefined as Positivity Rates. Throughout the entire revised manuscript, both the DLNM and LSTM pipelines have been completely re-analyzed using influenza positivity rates (laboratory-confirmed cases ÷ total ILI specimens tested) as the primary outcome, rather than raw case counts. The positivity rate is the WHOrecommended metric

for sentinel surveillance precisely because it mathematically standardizes the denominator, normalizing the raw testing volume variability. Its adoption effectively neutralizes the circular-reasoning concern raised by you. All predictive models, Figures 4, 5, and 6, and all primary metrics in Table 2 now reflect positivity-based estimates.

(2) Weekly detection volume integrated as a dynamic LSTM covariate. Even after positivity-rate normalization, residual testing-intensity effects can persist (e.g., if testing patterns shift among demographic subgroups with systematically different positivity profiles). To capture this, our revised LSTM network incorporated weekly detection volumes as an explicit dynamic input feature. As detailed in our covariate-ablation sensitivity analysis (Table 2), removing the testing-volume covariate drastically degraded the forecasting accuracy for 2023, increasing the Mean Absolute Error (MAE) by 22.2% for influenza A and an astounding 100% for influenza B. This indicates that the model's predictions are not driven solely by meteorological inputs but are appropriately and dynamically conditioned on the surveillance context.

(3) SHAP quantification of surveillance bias. By incorporating testing volume into the LSTM, we made the surveillance-bias concern concretely visible and quantifiable. As shown in our new SHAP analysis (Figure 6E-H), weekly_detection emerged as the second most impactful predictor of the positivity rate for influenza B, and, when examining the influenza A results, weekly_detection likewise ranked fifth among the most impactful predictors. This proves that the deep learning algorithm autonomously recognized the profound impact of testing intensity and actively utilized it to dynamically adjust and calibrate its epidemiological forecasts.

(4) Decoupling "Circular Reasoning" via DLNM Sensitivity Analysis. To definitively prove that our model does not merely "confirm historical opinions through increased testing," we conducted a targeted DLNM sensitivity analysis (detailed in the new Additional file 4 and Figures S3–S4). We compared DLNM exposure-response curves predicting positivity rates with and without adjusting for weekly testing volumes. The results were striking: the non-linear exposure-response curves and extreme-weather lag patterns remained highly consistent across both models. This empirical stability proves that the identified meteorological drivers represent intrinsic biological/environmental triggers of viral transmission, independent of fluctuating surveillance intensity.

We are deeply grateful that this critique prompted such substantial methodological refinement. The revised analyses are vastly more robust and epidemiologically interpretable than the original case-count-based version. We invite you to review the recently included sentences in the relevant subsections as specified above.

"To rigorously address the inherent confounding effects of "surveillance intensity bias", where fluctuations in raw case counts may merely reflect transient surges in clinical testing capacity rather than true community transmission, the primary outcome metric for all DLNM modeling was mathematically defined as the influenza positive rate, with meteorological factors and weekly detection volume serving as independent variables. The formula for calculating the daily influenza positivity rate is as follows:

$$R(\%) = \frac{I}{N} \times 100\%$$

where *R* represents the daily influenza positivity rate, *I* represents the number of daily influenza positive cases, and N denotes the total number of daily influenza tests performed. By adopting this WHO-recommended surveillance metric, our our analytical framework explicitly standardizes the epidemiological denominator. This mathematical normalization effectively neutralizes the severe surveillance intensity bias caused by dramatic testing

volume surges during the COVID-19 pandemic, ensuring that our models capture intrinsic viral transmissibility rather than artificial fluctuations in healthcare-seeking behavior or diagnostic capacity." (Methods, page 16)

"Furthermore, to evaluate the robustness of our findings against testing intensity, we conducted a targeted sensitivity analysis by reconstructing the DLNM models without the "weekly detection volumes" covariate. We then compared the non-linear cumulative risks and extreme weather lag effects between these ablated models and the original fully adjusted models. This allowed us to determine whether the identified climate-transmission associations were stable and biologically intrinsic, or merely artifacts of weather-correlated testing behaviors (Supplementary Information Additional file 4)." (Methods, page 17-18)

"To assess the contribution of pandemic-related confounding variables, we performed targeted sensitivity analyses on the LSTM networks. Specifically, we sequentially removed covariates of mask-wearing stringency indices and weekly detection volumes from the input features while maintaining identical Bayesian-optimized hyperparameters. The performance of these ablated models was evaluated using MAE, RMSE, MAPE, and SMAPE, allowing us to quantify the exact necessity of incorporating testing and behavioral covariates in forecasting models during periods of epidemiological non-stationarity." (Methods, page 22)

"Crucially, this LSTM stage explicitly incorporates weekly detection volumes, maskwearing stringency indices, non-local population proportion, and DOW effects alongside meteorological inputs, and uses influenza positivity rates rather than absolute case counts as the modeling endpoint. Together, these design choices are intended to mitigate, rather than fully eliminate, the surveillance-related biases that can distort count-based forecasting during periods of fluctuating testing intensity." (Discussion, page 40)

"This study has several limitations that should be considered when interpreting the findings. Firstly, our primary analysis relies on influenza surveillance data collected from seven sentinel hospitals in Putian, which inherently captures only a fraction of all influenza cases occurring in the broader community. Although employing positivity rates as the primary outcome substantially mitigates the surveillance bias inherent in count-based analyses, residual selection effects may persist if testing patterns shift differentially across demographic subgroups with systematically divergent positivity profiles. Our inclusion of weekly detection volumes as an explicit LSTM covariate, complemented by parallel DLNM sensitivity analyses validating the independence of meteorological effects from testing volumes, were designed to characterize and partially account for this residual bias, but cannot fully eliminate it. Consequently, positivity rates likely underestimate the true burden of community influenza infection. Although the surveillance infrastructure in Putian remained uniform and uninterrupted throughout the 2018–2023 timeline, mitigating measurement bias and temporal confounding risks, the transferability of our findings to other global subtropical regions with different socioeconomic structures, healthcare systems, or population behaviors requires cautious, region-specific calibration. Accordingly, our framework should be strictly interpreted as a high-fidelity tool designed to forecast the observable public health surveillance signal, which is the most operationally relevant target for public health agencies, rather than for estimating unobserved, absolute community disease burden." (Discussion, page 43-44)

"This limitation was particularly exacerbated by the profound epidemiological disruptions during the COVID-19 pandemic, where raw numbers of confirmed cases became heavily confounded by surveillance intensity (i.e., fluctuating testing volumes) rather than solely reflecting underlying viral transmission. Our adoption of influenza positivity rates as the primary modeling endpoint and incorporation of weekly detection volumes, face-covering stringency, and non-local population proportion as dynamic covariates, rigorously mitigated these aggregate-level biases and linked our methodological design directly to the forecasting outcomes. As unequivocally demonstrated by our covariate-ablation sensitivity analyses

(Table 2), failing to account for mask mandates and testing volumes leads to severe, mathematically predictable deviations in absolute forecasting accuracy. Furthermore, while daily case counts in a single city can occasionally be small, sporadic, and driven by external importations, our incorporation of non-local population proportion effectively adjusted for these localized importation risks. Consequently, although our findings characterize population-level associations between meteorological factors and influenza activity, they should not be interpreted as evidence of micro-level causal mechanisms at the individual patient level. Ultimately, the DLNM-LSTM framework's robust performance across the non-stationary transition out of NPI policies highlights the absolute necessity of integrating behavioral and virological baseline metrics into future climate-driven predictive surveillance systems." (Discussion, page 45-46)

We are deeply grateful that this critique prompted such substantial methodological refinement. The revised analyses are vastly more robust and epidemiologically interpretable than the original case-count-based version.

> *(1) Although the authors note a correlation between influenza and the weather factors. The authors do not discuss some of the high correlations between weather factors (e.g., solar radiation and UV index). Because of the many weather factors, those plots are hard to parse.*

We sincerely appreciate your constructive feedback regarding the visual clarity of the correlation plots and the methodological implications of highly correlated meteorological variables. We agree that the original 11 × 11 scatterplot matrix was visually overwhelming and could obscure the statistical implications of highly correlated features (such as solar radiation and the UV index, Pearson's $r > 0.9$).

To address this, we have taken two specific revisions:

(1) Improved Visualization (Revised Supplementary Figure S2): To make complex relationships easier to parse, we have completely redesigned Supplementary Figure S2. It is now logically partitioned into two distinct visual components:

- Panel A features a high-contrast Pearson correlation heatmap, where color intensity and statistical significance asterisks allow for rapid, intuitive identification of highly correlated pairs (such as solar radiation and UV index).

- Panel B retains the pairwise scatterplots (lower-left) and density distribution curves (diagonal) to facilitate the detailed visual inspection of non-linear trends, data skewness, and potential anomalies.

(2) Methodological Defense on Potential Collinearity: We did not arbitrarily eliminate these highly correlated variables because our two-stage modeling approach inherently mitigates multicollinearity risks associated with potential collinearity:

- For the DLNM analysis: To prevent coefficient instability caused by multicollinearity, all DLNM lagged analyses strictly employed univariate exposure-response models for meteorological factors (i.e., evaluating the relationship between a single meteorological factor and influenza positivity rates at a time, while controlling for long-term temporal trends and including other covariates as fixed terms). Consequently, the independent effect sizes and lag structures derived from the DLNMs are completely unaffected by inter-variable correlations.

- For the LSTM network: Unlike traditional multiple linear regression (or ARIMA) where potential collinearity inflates standard errors and destabilizes coefficients, Deep Learning architectures (LSTM) are natively robust to redundant features. The network's non-linear activation functions and gating mechanisms naturally weight overlapping signals during the

optimization process, effectively using redundant variables as a form of algorithmic regularization without compromising predictive stability.

We have expanded the Statistical Analysis and Discussion sections of the manuscript to explicitly articulate this rationale, ensuring maximum methodological transparency.

We invite you to review the recently included sentences in the relevant subsections as specified above.

“Prior to analytical modeling, we assessed the pairwise associations and potential collinearity among all meteorological variables using a comprehensive correlation heatmap and scatterplot matrix (Supplementary Figure S2). Notably, certain variables, such as solar radiation and the UV index, exhibited high positive correlations. To avoid coefficient instability typically caused by multicollinearity in regression models, all DLNM analyses adopted a univariate approach for meteorological factors, sequentially evaluating the nonlinear and lagged effects of individual meteorological predictors while adjusting for time trends and including other fixed covariates. Conversely, all variables were retained during the LSTM forecasting phase, as the non-linear gating architecture of recurrent neural networks inherently exhibits robust regularization against potential collinearity among input features.” (Results, page 26)

“Furthermore, extensive environmental inputs inevitably introduce severe collinearity, such as the strongly correlated solar radiation and UV index. While traditional multivariate models are highly vulnerable to such overlapping variances, the recurrent, weighted representation learned by the LSTM is comparatively tolerant of such redundancy, allowing broader covariate integration than in previous efforts.” (Discussion, page 40)

We hope that our responses and revisions will meet your expectations and demonstrate our dedication to improving the scientific quality of this study.

> *(2) The authors do not actually compare the results of both methods and what the LSTM adds.*

We sincerely appreciate your perceptive critique. We agree that our initial manuscript lacked a sufficiently rigorous, side-by-side comparison, and more importantly, it failed to adequately articulate why the LSTM architecture succeeds where classical statistical baselines fail.

To rectify this, we have comprehensively overhauled the comparative analysis. We formulated our baseline as a multivariate ARIMA model, supplying it with the exact same multidimensional covariate matrix as the LSTM (including meteorological variables, mask-wearing indices, and weekly testing volumes), focusing on three dimensions:

(1) Direct Quantitative Comparison: Following the restructuring of our outcome variable to the Influenza Positivity Rate, we re-evaluated both models using identical training (2018–2022) and validation (2023) datasets. The LSTM consistently and substantially outperformed the ARIMA model across all metrics. For Influenza A, the LSTM achieved an MAE of 0.009 and RMSE of 0.035 (vs. ARIMA: MAE 0.136, RMSE 0.238). For Influenza B, the LSTM yielded an MAE of 0.002 and RMSE of 0.011 (vs. ARIMA: MAE 0.049, RMSE 0.057). We have updated Table 2 and the corresponding Results section to explicitly present these side-by-side comparisons.

(2) Visualizing “Where” the LSTM Outperforms: We have updated Supplementary Figure S3 to map the ARIMA predictions for the 2023 positivity rate, allowing for a direct visual comparison with the LSTM predictions in Figure 6 (A, B). The visualizations explicitly reveal the ARIMA model’s fundamental limitation: it tends to predict relatively flat or conservatively smoothed values, failing entirely to capture the extreme, explosive non-linear peaks of the 2023 viral rebound. Conversely, the LSTM network accurately tracks these sudden epidemic phase transitions.

(3) Articulating "What the LSTM Adds" (Discussion Expansion): We have significantly expanded the Discussion section to intellectually articulate why the LSTM succeeds where ARIMA fails. To ensure a strictly fair methodological comparison, we formulated our baseline as a multivariate ARIMA model (ARIMAX), supplying it with the exact same multidimensional covariate matrix as the LSTM (including meteorological variables, mask-wearing indices, and weekly testing volumes). Therefore, the LSTM's superior performance is not due to information asymmetry (i.e., it did not "see" more variables), but stems directly from its algorithmic architecture. The consistent superiority of the LSTM over both the linear sequential baseline (ARIMA) and the non-linear non-sequential baseline (XGBoost) isolates the recurrent gated architecture itself as the source of the predictive gain. This advantage arises from four architectural properties intrinsic to recurrent gated networks yet absent in both tree ensembles and linear autoregressive models:

a) Sensitivity to temporal ordering, which decision-tree splits and linear regressors cannot natively encode;

b) Gated propagation of long-range dependencies through forget–input–output mechanisms;

c) Explicit accommodation of serial autocorrelation, violated by the i.i.d. assumptions underlying gradient-boosted trees;

d) Paradigmatic comparability with sequential statistical models: the joint failure of ARIMA (linear, sequential) and XGBoost (non-linear, non-sequential) isolates recurrent sequential memory, rather than non-linearity per se, as the critical feature for forecasting under pandemic-era non-stationarity.

We emphasize that this interpretation applies specifically to the present non-stationary epidemiological forecasting task and does not constitute a general dismissal of gradient-boosted ensembles, which retain competitive performance across many structured prediction domains.

We invite you to review the recently revised table, figures and sentences in the relevant subsections as specified above.

"The LSTM networks accurately captured both the timing and magnitude of these nonlinear epidemic surges, including the two outbreak peaks of influenza A during February-March and November-December of 2023, as well as the peak of influenza B in November-December, demonstrating good predictive performance. The predictive performance of the LSTM networks was quantitatively assessed using metrics such as MAE, RMSE, MAPE, and SMAPE. For influenza A, the MAE was 0.009, RMSE was 0.035, MAPE was 0.158, and SMAPE was 0.521; for influenza B, the MAE was 0.002, RMSE was 0.011, MAPE was 0.17, and SMAPE was 0.484 (Table 2)." (Results, page 31)

"To benchmark the predictive value added by the LSTM architecture, we constructed two baseline models using identical training (2018–2022) and validation (2023) positivity-rate datasets, inclusive of all contextual covariates: the multivariate ARIMA model and the XGBoost gradient-boosting model. The LSTM model demonstrated decisive superiority over both baselines across all evaluated metrics (Supplementary Table S3). For influenza A, the LSTM achieved an MAE of 0.009 and SMAPE of 0.521, compared with the ARIMA's MAE of 0.136 (SMAPE 1.212) and XGBoost's MAE of 0.138 (SMAPE 1.081), representing approximately 15-fold error reductions relative to both benchmarks. For influenza B, the LSTM yielded an MAE of 0.002 (SMAPE 0.484), compared with 0.049 (SMAPE 0.810) for ARIMA and 0.070 (SMAPE 0.737) for XGBoost, approximately 25- to 35-fold reductions. Notably, ARIMA and XGBoost produced errors of comparable magnitude despite their disparate assumptions regarding linearity, suggesting that the LSTM's advantage derives not from non-linear

modeling capacity per se, but from architectural properties specific to recurrent sequential processing.

Beyond global error metrics, visual comparison of forecasting trajectories (Figure 6; Supplementary Figures S5 and S6) reveals critical behavioral disparities. For influenza A, both the linear ARIMA model and the non-linear XGBoost model generated conservatively smoothed forecasts that entirely failed to capture the sudden, explosive peaks of the 2023 post-restriction viral rebound. For influenza B, ARIMA produced continuous spurious fluctuations during non-epidemic periods (when true positivity was near zero), likely overreacting to covariate variations, while XGBoost generated largely flat trajectories that missed the mid-year outbreak peak. In stark contrast, the LSTM's recurrent gating mechanisms successfully filtered out covariate noise during low-transmission periods while accurately tracking extreme epidemiological phase transitions, demonstrating a qualitative advantage in handling non-stationary regime shifts." (Results, page 33-34)

"To rigorously isolate the predictive contribution attributable to the LSTM's recurrent architecture, we benchmarked it against two covariate-matched baselines representing distinct methodological paradigms: the multivariate ARIMA model and the XGBoost gradient-boosting ensemble. This design controls simultaneously for linearity (ARIMA→LSTM contrast) and for non-linearity without recurrent memory (XGBoost→LSTM contrast), allowing us to attribute observed performance gains to specific architectural inductive biases rather than to informational asymmetry or model non-linearity in general. The LSTM substantially outperformed both baselines on the 2023 validation window. For influenza A, it achieved an MAE of 0.009, compared with 0.136 for ARIMA and 0.138 for XGBoost, approximately 15-fold reductions. For influenza B, the LSTM yielded an MAE of 0.002, versus 0.049 for ARIMA and 0.070 for XGBoost, 25- to 35-fold reductions. Critically, ARIMA and XGBoost produced errors of comparable magnitude despite their disparate assumptions regarding linearity, and both systematically under-predicted the explosive 2023 post-NPI rebound for influenza A while generating flat or spurious trajectories for influenza B (Supplementary Figures S5–S6). That parallel failure indicates the LSTM's advantage under pandemic-era non-stationarity derives not from non-linearity per se, but from four architectural properties intrinsic to recurrent gated networks yet absent in tree ensembles: (i) threshold-based, sequence-insensitive splits fail to encode present–past dynamics; (ii) XGBoost lacks forget–input–output gates to propagate and re-weight historical states across time lags; (iii) gradient-boosted trees assume near-independence, contradicting the pronounced temporal autocorrelation in epidemiological time series; (iv) ARIMA and LSTM are sequential models with different functional forms, whereas XGBoost is non-sequential. The joint failure of ARIMA and XGBoost, despite differing non-linear treatment, isolates recurrent sequential memory as the critical architectural feature for forecasting under non-stationarity. We emphasize that this interpretation applies specifically to the present non-stationary influenza forecasting task and does not constitute a general dismissal of gradient-boosted ensembles, which retain state-of-the-art performance across many structured prediction domains. Rather, it highlights that for surveillance time series exhibiting pronounced temporal dependencies and abrupt regime shifts, such as the 2023 post-NPI rebound, explicit sequential memory becomes functionally essential. This mechanistic reading, together with the LSTM's comparative edge over previously reported ARIMA-based (Li et al. 2024) and LSTM-based influenza prediction models (Zhu et al. 2022), positions our framework as a substantive methodological advance in predictive modeling for climate-sensitive diseases." (Discussion, page 41-42)

Results

"The MAE for the covariate-adjusted ARIMA model of influenza A was 0.136, RMSE was 0.238, MAPE was 1.115, and SMAPE was 1.212. For influenza B, the MAE was 0.049, RMSE was 0.057, MAPE was 1.426, and SMAPE was 0.810. Overall, the predictive performance of the ARIMA model was substantially inferior to that of the Bayesian optimized LSTM. As illustrated in

Figure S5, the linear model struggled significantly with the non-stationary dynamics of the 2023 viral rebound. It either failed entirely to capture the extreme, explosive non-linear peaks (as seen in Influenza A) or generated continuous spurious predictions during zero-case periods due to mechanical linear reactions to covariate inputs (as seen in Influenza B)." (Supplementary information, page 16)"

Results

"The XGBoost model produced substantially higher forecasting errors than the LSTM across both influenza subtypes (Table S3; Figure S6). For Influenza A, XGBoost yielded MAE = 0.1379, RMSE = 0.2161, MAPE = 1.8195, and SMAPE = 1.0805, errors of a magnitude broadly comparable to the multivariate ARIMA baseline (MAE = 0.136) and approximately 15-fold higher than the LSTM (MAE = 0.009). For Influenza B, XGBoost achieved MAE = 0.0704, RMSE = 0.0739, MAPE = 1.9172, and SMAPE = 0.7371, exceeding both the ARIMA benchmark (MAE = 0.049) and the LSTM (MAE = 0.002) by approximately 35-fold relative to the latter." (Supplementary information, page 19)

We believe these additions provide a much more rigorous and academically satisfying comparative analysis.

> *(3) The methods are long and meandering. They could be cleaned up and shortened. E.g., there is no need for 30 lines on PCR testing; the study area should come before the study design. The authors discuss similar elements in multiple places; this whole section can be shortened considerably without affecting the content.*

We sincerely appreciate the reviewer's editorial guidance. We agree that the initial Methods section was somewhat disjointed and unnecessarily verbose, reflecting iterations from previous drafts. We have completely restructured and aggressively streamlined this section to ensure a logical and concise flow:

- Structural Reorganization: As suggested, we have moved the "Study area" section to the very beginning of the Methods, providing the geographical and climatic context before detailing the study design and cohort.

- Consolidation of Redundancies: We have merged the fragmented descriptions regarding ethical approvals, data anonymization, and sentinel hospital protocols into a single, cohesive "Study design and cohort" subsection.

- Condensing Laboratory Protocols: We completely agree that 30 lines on standard PCR testing in the main text are unnecessary. We have condensed the "Sample collection and pathogen typing" section into a brief, 4-line summary explicitly stating the use of commercial assays (Da'an Gene Co., Ltd) and strict adherence to China CDC guidelines. The highly technical nuances (e.g., RNA extraction integrity, spectrophotometer ratios, and precise PCR amplification thresholds) have been relocated to the Supplementary Information for interested readers.

We invite you to review the recent revisions in the Methods subsection as specified above.

"Sample collection and pathogen typing

Respiratory specimens (nasopharyngeal swabs) were collected from ILI patients during their initial visit, prior to treatment, and stored at 4°C in viral transport medium. All samples were delivered to the Putian CDC laboratory within 24 hours of collection. Nucleic acid extraction and one-step real-time fluorescent RT-PCR for influenza A/B subtyping (H1N1, H3N2, Victoria, and Yamagata lineages) were executed within 24 hours upon sample arrival. All assays utilized commercial diagnostic kits (supplied by Da'an Gene Co., Ltd., Guangzhou, China) and were processed in a Biosafety Level 2 (BSL-2) laboratory, strictly adhering to the

manufacturer's instructions and China CDC's standardized protocols. Detailed laboratory procedures, including RNA integrity parameters and PCR amplification thresholds, are comprehensively documented in the Supplementary Information." (Methods, page 13)

These revisions have significantly improved the readability of the manuscript without sacrificing methodological transparency.

> *(4) How reliable is the "Our Word in Data" website for subnational coverage of restrictions? Some of the authors are from Putian and should be able to confirm the accuracy for both studied areas.*

We are very grateful for this insightful question. You astutely identify a common challenge in geospatial epidemiology in China: the systemic lack of publicly accessible, standardized, daily non-pharmaceutical intervention (NPI) datasets at the municipal (subnational) level. Local CDC policy records are typically maintained as internal administrative documents without standardized time-series data interfaces.

Given this limitation, we utilized the national Our World in Data (OWID) stringency index as a proxy. To directly address the reviewer's excellent point regarding local verification, the co-authors of this study—who are frontline epidemiologists stationed at the Putian CDC and who directed the local pandemic response, including conducted a rigorous, retrospective cross-validation of the OWID index against local realities.

Our local experts confirmed a high degree of fidelity between the OWID 0–4 scale and the actual policies enforced in Putian and Sanming:

- 2020 (Initial Outbreak): Both cities enforced "mandatory face coverings outside the home at all times" (OWID Level 4), aligning perfectly with the dataset.

- 2021–2022 (Normalized Control): Policies shifted to "required in all shared/public spaces" (OWID Level 3), which accurately reflects the local mandates required for public transit, schools, and commercial venues.

- 2023 (Post-Pandemic Shift): Following the national policy pivot, mandates were downgraded to "recommended" (OWID Level 1), perfectly mirroring local ground truths.

Therefore, while OWID provides a national-level index, our local CDC authors have empirically verified that its temporal variations accurately capture the intensity of behavioral restrictions experienced by the populations in our specific study areas. We have incorporated a concise statement regarding this expert validation into the revised Methods section.

"While OWID stringency indices represent national-level policy, publicly accessible and standardized daily NPI datasets at the municipal level are currently unavailable in China. To ensure the spatial validity of these indices, co-authors from Putian CDC, who actively managed the local epidemic response, conducted a rigorous cross-validation. Our local public health experts confirmed that temporal fluctuations of OWID indices (ranging from Level 4 strict mandates in 2020 to Level 1 recommendations in 2023) exhibited high fidelity with the actual, on-the-ground enforcement of NPIs in both Putian and Sanming. Thus, OWID stringency indices serve as a highly reliable contextual proxy for local social contact restrictions." (Methods, page 15)

We hope that our responses and the additional statement will meet your expectations.

> *(5) Figure 2A is hard to parse; it would make more sense to plot these as line plots (y=count, x=month).*

We completely agree with you. The original heatmap visualization obscured the temporal dynamics of the distinct influenza subtypes. We have entirely redrawn Figure 2A as multi-line plots mapping the monthly incidence trajectories of Influenza A (H1N1, H3N2) and Influenza B (Victoria, Yamagata) across the six-year study period. This new visualization (included in the revised manuscript) vastly improves readability and explicitly highlights the distinct phase shifts and interruptions caused by the pandemic.

> ***Reviewer #1 (Recommendations for the authors):***
>
> *(1) Figure 3 is hard to parse. The flu counts are repeated in every figure, which makes them look very similar. I would recommend that the authors pick 1 or 2 as subpanels for the main paper and put the rest in the supplement.*

We sincerely appreciate your feedback. We completely agree that the original 3x3 square layout severely compressed the x-axis, making the daily temporal fluctuations difficult to parse and causing the flu curves to look visually redundant.

To resolve this visual clutter without losing valuable environmental context in the main text, we adopted a highly effective structural solution simultaneously suggested by Reviewer #2. We have completely redrawn Figure 3 into an 8x1 vertically stacked format with a single, shared continuous x-axis across the entire study period. This extended aspect ratio vastly expands the timeline, dramatically revealing the highly distinct, daily microfluctuations of each meteorological factor alongside the epidemiological curves. We believe this new layout completely resolves the parsing difficulty you rightly pointed out. Given that a substantial portion of our readership relies heavily on the main text figures for immediate epidemiological context, retaining these cleanly formatted panels in the main manuscript maximizes the paper's scientific impact. We hope you find this redesigned visualization satisfactory.

> *(2) Using a thousand-separator throughout will make the manuscript more readable.*

We completely agree. We have meticulously applied thousand-separators to all relevant numerical values (e.g., 20,488; 17,333) throughout the revised manuscript to enhance readability.

> *(3) Line 556 "quantified calculated", pick 1 word.*

We sincerely apologize for this typographical oversight resulting from the drafting process. However, the original sentence that led to the duplicated phrasing you highlighted has been removed, as we had already undertaken a comprehensive revision of the relevant material in that subsection in response to your earlier remarks. We invite you to review the newly substituted paragraph below.

"Prior to analytical modeling, we assessed the pairwise associations and potential collinearity among all meteorological variables using a comprehensive correlation heatmap and scatterplot matrix (Supplementary Figure S2). Notably, certain variables, such as solar radiation and the UV index, exhibited high positive correlations. To avoid coefficient instability typically caused by multicollinearity in regression models, all DLNM analyses adopted a univariate approach for meteorological factors, sequentially evaluating the nonlinear and lagged effects of individual meteorological predictors while adjusting for time trends and including other fixed covariates. Conversely, all variables were retained during the LSTM forecasting phase, as the non-linear gating architecture of recurrent neural networks inherently exhibits robust regularization against potential collinearity among input features." (Results, page 26)

> ***Reviewer #2 (Public review):***

> *Summary:*
>
> *The study aimed to assess the associations between meteorological drivers and influenza is important although not new. The authors used only 6 years of surveillance data and deep learning models, combining distributed lag non-linear models (DLNM) with Bayesian optimized LSTM neural networks for predictive modeling. The key interest in this area is to explore the subtropical locations, where influenza is less common and circulates year round. The authors further claimed that such an association could be able to provide an early warning in the community. In this direction, the current manuscript has several scopes of improvements and clarification of the claims, as I list here.*
>
> *Strengths:*
>
> *Study design based on a prospective cohort to analyse the data for retrospective outcomes.*

We sincerely thank you for the careful and constructive evaluation of our manuscript, and in particular for recognising the value of our prospective surveillance design and the importance of investigating influenza–meteorological associations in subtropical settings where year-round circulation patterns differ substantively from those in temperate regions. We are grateful that you have identified four specific dimensions in which the manuscript can be strengthened, rationale clarity, methodological/data-integration transparency, validation reporting, and the calibration of the "early warning" claim. We address each of these four points in detail below, and we have undertaken substantive revisions to the manuscript in response.

> *Weaknesses:*
>
> *(1) The rationale of the study is not clearly stated.*

We sincerely thank you for this incisive observation. We agree that the original Introduction did not adequately articulate the study's rationale, specifically, the causal chain linking public-health need, existing methodological limitations, and the incremental contribution of our integrated DLNM-plus-LSTM framework. The Introduction has been substantively rewritten to make this rationale explicit, structured around four logical pillars:

- Disease burden grounding. We have added quantitative evidence on the global burden of seasonal influenza, such as annual mortality estimates, drawing on solid epidemiological sources, to establish the public-health magnitude that motivates the study.

- Subtropical-specific knowledge gap. We articulated the distinctive epidemiological challenges of subtropical influenza transmission, including year-round circulation patterns, complex non-linear meteorological associations, and lag-structured exposure-response relationships, that fundamentally differentiate subtropical contexts from temperate epidemiological settings where most existing research has been conducted. This articulation directly motivates our adoption of distributed lag non-linear models (DLNM) as the appropriate analytical framework for capturing these complex non-linear and lag-structured associations.

- Methodological gap and incremental contribution. We now position our integrated framework against three specific gaps in the existing literature: (a) studies using DLNM alone characterize lag-distributed exposure–response relationships but lack forecasting capability; (b) studies using LSTM alone provide forecasts but typically do not incorporate Bayesian hyperparameter optimization, do not stratify by influenza subtype, and do not account for COVID-19-era non-pharmaceutical interventions; (c) no existing study, to our knowledge, integrates DLNM-based mechanistic interpretation with Bayesian-optimized, subtype-specific

LSTM forecasting in a subtropical Chinese setting under pandemic-perturbed surveillance conditions. Our study is positioned to fill this specific gap.

- Adequacy of the six-year data window. We additionally address your implicit concern regarding study duration. While six years (2018–2023) is shorter than some long-horizon influenza time-series studies, this window was deliberately selected because it brackets a uniquely informative epidemiological transition: two prepandemic baseline years (2018–2019), three Non-Pharmaceutical Interventions (NPI)-suppressed years (2020–2022), and one post-suppression rebound year (2023). This structure allows the model to learn from a structural break that a longer but earlier-only series could not provide. We have made this argument explicit in the revised Introduction.

- The dual-model rationale. We clarify that our core rationale is to bridge this gap. By utilizing DLNM to uncover the underlying environmental biological triggers and subsequently employing the Bayesian-optimized LSTM network, uniquely upgraded to ingest epidemiological context (mask-wearing stringency index, testing volumes, day and week, viral importation), we provide a comprehensive framework that achieves both mechanistic insight and operational forecasting agility. We have substantially rewritten the Introduction to reflect this explicit storyline.

We close the revised Introduction by explicitly framing the study's translational endpoint: providing a methodologically integrated framework (DLNM for mechanistic interpretation; Bayesian-optimised LSTM for forecasting) to support climate-informed influenza preparedness in subtropical settings. We have, however, calibrated the language used to describe this endpoint (see our response to Weakness #4) to avoid overstating the study's operational readiness as an early-warning tool.

We have substantially rewritten the Introduction to reflect this sharpened rationale. We invite you to read the whole section of the Introduction in the revised manuscript.

> *(2) Several issues with methodological and data integration should be clarified.*

We sincerely appreciate your identification of methodological and data integration ambiguities in the original manuscript. We have substantively addressed this concern through three categories of clarifying revisions: (i) explicit articulation of the covariate framework, (ii) clarification of the analytical relationship between DLNM and LSTM components, and (iii) detailed specification of data sources and quality control procedures:

- Clarification 1: Comprehensive covariate framework specification. We have explicitly articulated the complete covariate framework integrated into both DLNM and LSTM analyses. Beyond the meteorological factors (mean temperature, maximum temperature, minimum temperature, diurnal temperature range, relative humidity, atmospheric pressure, precipitation, sunshine duration), our analytical framework systematically incorporates: (a) influenza positivity rates as the primary outcome variable (replacing raw case counts to mitigate surveillance intensity bias, as detailed in our response to Reviewer 1's Public Review Comment 2); (b) weekly testing volumes as an explicit covariate to control for residual surveillance-intensity variations; (c) mask-wearing stringency indices to capture pandemic-era public health intervention effects; (d) day-of-week indicators distinguishing weekdays from weekends to control for healthcare-seeking behavioral cycles; and (e) nonlocal population proportion to account for population mobility-related transmission dynamics.

- Clarification 2: Articulation of the DLNM-LSTM analytical relationship. We have explicitly clarified the complementary analytical roles of DLNM and LSTM within our integrated framework, addressing potential confusion regarding whether these methods serve redundant or complementary functions.

- Clarification 3: Data source specification and quality control documentation. We have substantively expanded the data source specification and quality control documentation to ensure full methodological transparency.

We have revised the "Study design" subsection in the Methods to transparently outline how these multifaceted data streams were temporally aligned and fed into the dual-model architecture. We invite you to review these rewritten paragraphs.

"The study spanned January 1, 2018, to December 31, 2023, integrating four categories of data sources: (i) ILI and laboratory-confirmed cases from seven influenza sentinel hospitals across Putian's urban and rural areas, ensuring representative coverage of diverse healthcare-seeking populations; (ii) daily meteorological data; (iii) COVID-19 associated public health intervention indicators (mask-wearing stringency indices) recorded from January 2020 onwards; and (iv) demographic mobility indicators (non-local population proportion) obtained from ILI consultation records. To ensure consistency between meteorological measurements and influenza incidence records across all data sources, we applied rigorous quality control and pre-processing procedures, including: temporal alignment of all data streams to a unified daily resolution; missing value imputation using temporally adjacent observations for sporadic gaps (<5% of records); cross-validation of laboratory-confirmed cases against ILI consultation records to identify and resolve coding inconsistencies; and standardization of meteorological measurements against the regional monitoring network's established calibration protocols. Final datasets underwent independent verification by two co-investigators to ensure analytical reliability." (Methods, page 11)

"DLNM was first constructed to screen meteorological factors and other covariates with substantial influence on influenza seasonality. Subsequently, an LSTM neural network was developed within the same covariate system. The integrated DLNM–LSTM framework was employed as methodologically complementary rather than redundant components, leveraging the distinctive strengths of each approach to address different analytical objectives within a unified investigation. Specifically, DLNM models characterize the non-linear exposure–lag–response relationships between meteorological factors and influenza risk, providing biologically interpretable insights into the temporal structure of weather-influenza associations and identifying meteorological factors with statistically and clinically significant effects on influenza dynamics. Building upon this DLNM-derived foundation, the LSTM network constructs a time-series forecasting tool within an identical covariate framework, evaluating predictive capability for influenza transmission trends. Beyond meteorological factors, our DLNM-LSTM framework systematically incorporated influenza positivity rates as the primary outcome variable, weekly detection volumes, mask-wearing stringency index (indicator of NPIs during the COVID-19 pandemic), day of the week (DOW, distinguishing weekdays from weekends), and non-local population proportion (defined as the ratio of the number of individuals whose reported residential district at the time of testing lies outside Putian city to the total number of tests) as covariates within both DLNM and LSTM modeling pipelines. This comprehensive covariate framework ensures that observed meteorological associations are estimated after controlling for surveillance intensity, public health intervention status, behavioral healthcare-seeking cycles, and population mobility patterns." (Methods, page 11-12)

> *(3) Validation of the models is not presented clearly.*

We sincerely appreciate your identification of insufficient clarity in the validation framework presentation. We acknowledge that the original manuscript inadequately articulated the multi-tiered validation architecture underlying our analytical framework. We have substantively expanded the validation framework documentation through three categories of clarifications: (i) explicit articulation of the three-tier data partitioning architecture, (ii)

detailed specification of validation procedures across each tier, and (iii) systematic enumeration of validation evidence supporting each analytical conclusion.

- Clarification 1: Three-tier data partitioning architecture. Our validation framework employs a rigorously designed three-tier data partitioning architecture that addresses different validation objectives at each tier.

Tier 1: Internal training and validation (Putian, 2018-2022). We allocated the 2018-2022 Putian surveillance data as the primary training set, within which 10% of samples were further randomly partitioned as an internal validation subset for hyperparameter tuning and overfitting monitoring during LSTM training. Early stopping mechanisms were implemented to terminate training when internal validation loss plateaued, preventing overfitting to training-specific patterns.

Tier 2: Internal testing (Putian, 2023). We allocated the 2023 Putian surveillance data as the internal testing set, providing a temporally independent assessment of model predictive performance on data not utilized during training or hyperparameter optimization. The chronological partitioning preserves time series modeling validity by ensuring that all training data temporally precede testing data, avoiding data leakage that could artificially inflate performance estimates.

Tier 3: External validation (Sanming, 2023). We obtained surveillance data from Sanming city for the period January 1, 2023, to December 31, 2023, matching the temporal coverage of the Putian internal testing set. This external validation set provides geographically independent assessment of model transferability across subtropical Chinese contexts, evaluating whether the Putian-derived model architecture generalizes to a different subtropical city sharing comparable climatic characteristics, influenza seasonality patterns, and public health intervention frameworks.

- Clarification 2: Unified model framework across all validation tiers. A critical methodological feature of our validation framework is that the identical Bayesian-optimized LSTM architecture trained on Putian 2018-2022 data was applied without modification across all three validation tiers. This unified framework approach is methodologically essential because: (a) it tests genuine model transferability rather than evaluating differently-tuned models at each tier, which would conflate validation with re-optimization; (b) it enables direct performance comparison across internal testing and external validation, isolating the marginal performance degradation attributable to geographic transfer; and (c) it aligns with operational deployment scenarios where a trained model must be applied to new contexts without re-training.

- Clarification 3: Validation evidence enumeration. The validation evidence supporting our analytical conclusions encompasses four complementary dimensions:

Predictive performance metrics: Across both influenza A and B, the Bayesian-optimized LSTM achieved low error metrics on the internal testing set (influenza A: MAE = 0.009, RMSE = 0.035, MAPE = 0.158, SMAPE = 0.521; influenza B: MAE = 0.002, RMSE = 0.011, MAPE = 0.170, SMAPE = 0.484), substantially outperforming ARIMA benchmark models (Supplementary Figure S3).

External validation: The Putian-derived LSTM successfully generalized to Sanming external validation data, with performance metrics maintaining comparable magnitudes to internal testing performance, substantiating model transferability across subtropical contexts.

Sensitivity analyses: We conducted systematic sensitivity analyses across both DLNM and LSTM components. DLNM sensitivity analyses (Supplementary Figures S4-S5) demonstrate substantial concordance in cumulative risk patterns and lag-specific extreme condition responses across models with and without weekly testing volume adjustment, substantiating

robustness of meteorological associations. LSTM sensitivity analyses (New Table 2) demonstrate that systematic covariate exclusion produces predictable and biologically plausible performance degradation patterns rather than artificially robust performance, confirming the absence of overfitting characteristics.

Interpretability verification: SHAP interpretability analysis (Figure 6E-H) substantiates that the LSTM autonomously identified epidemiologically plausible feature importance hierarchies, providing independent verification that model predictions reflect genuine biological signal recognition rather than data artifacts.

We invite you to review the corresponding manuscript clarifications.

“The LSTM network was trained on data from Putian corresponding to the four categories described above, with the time period 2018-2022, and Putian’s 2023 data serving as the internal validation set. To rigorously evaluate model transferability beyond the training context, data from Sanming city, a mountainous subtropical city exhibiting comparable climatic characteristics, influenza seasonality, and public health intervention frameworks to Putian, were acquired for the period January 1, 2023, to December 31, 2023, temporally aligned with the Putian internal validation set. This Sanming dataset constituted our external validation set, facilitating a geographically independent assessment of model generalization within subtropical Chinese environments. The predictive performance of the LSTM algorithm for influenza A/B prevalence in 2023 Putian data was benchmarked against a parallel multivariate ARIMA model with exogenous variables. To ensure a fair methodological comparison, this baseline model was supplied with the exact same meteorological and epidemiological covariate matrix as the LSTM.” (Methods, page 12-13)

“To respect the temporal dependence inherent in LSTM architectures and avoid data leakage, we adopted a strict chronological out-of-time (OOT) validation strategy. Time-series data from January 1, 2018, to December 31, 2022 (88.26% of the Putian dataset) were used for model training, with 10% reserved during Bayesian optimization as an internal validation subset for convergence monitoring and hyperparameter tuning only. Data from January 1, 2023, to December 31, 2023 (11.74%) were held out as a chronologically internal validation set for final performance evaluation, covering a complete annual cycle. External validation was conducted using concurrent 2023 data from Sanming city to assess spatial generalizability. To prevent distributional leakage, all normalization parameters were derived exclusively from the training set and consistently applied to the validation sets, with predictions subsequently transformed back to the original scale. First, we performed data normalization, a crucial step to ensure that training and test set data are compared on a unified scale. We normalized the training and test set data separately within the range [0, 1]. For the test set normalization, we used the maximum and minimum values from the training set as boundaries. This approach ensured consistency between the normalized test set data and the training set data. After the algorithm conducted predictions on the test set data, we performed denormalization to convert the predicted results back to the original data scale and rounded them to integers. These steps ensured that the final prediction results accurately and objectively reflected the LSTM’s performance in real-world scenarios and provided reliable data for subsequent calculation of evaluation metrics.” (Methods, page 18-19)

> *(4) The claim for providing tools for 'early warning' was not validated by analysis and results.*

We are grateful for this incisive critique, which identifies a critical mismatch between our research achievements and the terminology employed in the original manuscript. Upon careful re-examination, we acknowledge unreservedly that the original manuscript’s use of “early warning” terminology overstated our actual research contribution. Our research has constructed and validated a methodologically rigorous LSTM-based influenza forecasting framework demonstrating strong predictive performance and external transferability;

however, this constitutes a forecasting framework foundation rather than a fully-validated operational early warning tool ready for direct public health implementation.

We recognize that genuine early warning tools require additional validation dimensions that our current research does not yet comprehensively address. In response to this important critique, we have implemented three categories of substantive corrections:

- Correction 1: Comprehensive terminology revision throughout the manuscript. We have systematically revised “early warning” terminology throughout the manuscript, replacing it with more accurate descriptors that precisely characterize our actual research contribution. Specifically: “early warning system” has been revised to “forecasting framework” or “forecasting model”; “early warning tool” has been revised to “predictive modeling foundation”; and “early warning capability” has been revised to “predictive capability supporting future early warning system development”. These terminological refinements ensure that manuscript claims precisely correspond to demonstrated research achievements.

- Correction 2: Manuscript title revision. We have correspondingly revised the manuscript title to remove “early warning” terminology and accurately reflect the study’s actual contributions: Revised title: “Meteorological Drivers of Influenza A and B Positivity in a Subtropical Chinese City: A Six-Year Surveillance Study Integrating Distributed Lag Non-Linear Models and Deep Learning”. This revised title precisely articulates the study’s actual scope: characterization of meteorological drivers (DLNM contribution), focus on positivity rates (methodological refinement addressing surveillance bias), specification of subtropical context (geographic scope), six-year temporal coverage (data scope), and integration of DLNM and deep learning (methodological framework).

- Correction 3: Explicit articulation of forecasting framework versus operational early warning tool distinction. We have explicitly articulated the distinction between our current achievements and operational early warning tool requirements in both the Discussion and Conclusion sections, framing future research directions for operational early warning system development.

We invite you to review the corresponding manuscript clarifications.

“Importantly, however, this framework should be regarded as a methodological foundation for future operational developments rather than as a deployable early-warning system: routine use in public health practice would require prospective recalibration, integration with operational surveillance infrastructure, and additional validation beyond the scope of the present study.” (Discussion, page 43)

“In conclusion, this study elucidates the distinct, non-linear meteorological drivers of influenza A and B transmission in a subtropical Chinese urban setting through an integrated dual-stage DLNM-LSTM framework. By adopting influenza positivity rates as the primary outcome and integrating socio-behavioral covariates, including mask-wearing stringency indices, weekly detection volumes, and non-local population proportion indicators, our approach mitigates surveillance-related biases and accommodates pandemic-era nonstationarity. It shows lower forecast error than a covariate-matched ARIMA baseline and provides preliminary evidence of portability within southeastern subtropical China, combining the interpretability of distributed lag modeling with the flexibility of deep learning. The framework offers an interpretable, climate-informed methodological foundation for future operational surveillance developments in subtropical settings. (Discussion, page 47)

> ***Reviewer #2 (Recommendations for the authors):***
>
> *(1) The title is not data-driven in different contexts, including 'early warning'; I was expecting substantial analyses in this direction to assess the 'early warning' in the*

> *manuscript. But I hardly found them in the text, merely utter as the implication of understanding the associations between meteorological drivers and influenza in advance. I suggest either revising the title or clarifying the claim by providing significant evidence and its impact on the epidemic onset and intensity. Further, revise 'Subtropical China' as 'a Subtropical Chinese city', as the former one is not accounted under this study.*

We completely agree with this constructive feedback. As detailed in our response to your Public Review weakness (4), we acknowledge that claiming an operational "early warning" system requires extensive real-world feasibility and threshold validations that exceed the scope of our current time-series analysis. We understand that achieving early warning capabilities necessitates the completion of at least three additional validation dimensions, as listed below.

Dimension 1 - Threshold determination: Early warning systems require explicit thresholds defined by integrating predictions with established epidemiological thresholds, typically via ROC analysis to balance sensitivity and specificity for trigger activation; our framework provides predictions but does not define thresholds.

Dimension 2 - Deployment validation: Validation of warning timeliness, false alarm control, and integration with existing CDC surveillance architectures is required; our work does not assess these operational dimensions.

Dimension 3 - Robustness across heterogeneous scenarios: Early warning tools need systematic robustness testing across diverse social environments, public health policy contexts, extreme meteorological events, and co-circulation of emerging pathogens; our results cover 2018–2023 Putian-Sanming but not broader operational scenarios.

We have modified the title to "a Subtropical Chinese city" and systematically revised "early warning" terminology to "forecasting" throughout the manuscript. These revisions ensure accurate representation of the study's scope and contribution.

> *(2) There are several studies establishing the potential association between influenza and the climatic drivers in several locations across the globe. I couldn't find sufficient text on establishing the rationale of this study from the perspective of existing literature. This should clearly be uttered in the introduction section itself.*

We are grateful for this critique, which echoes your Public Review weakness (1). We fully agree that the original Introduction lacked a cohesive narrative connecting the existing literature to our specific methodological innovations.

To address this, we have comprehensively rewritten the Introduction section. The revised text now systematically establishes our rationale through a clear logical progression:

- Acknowledging existing studies on climatic drivers but highlighting the unique challenge of non-linear, year-round influenza transmission in subtropical regions.

- Identifying the methodological gap: existing models either use DLNM purely for retrospective explanation (lacking prediction) or employ deep learning (LSTM) purely for prediction (lacking epidemiological interpretability).

- Highlighting the critical failure of current literature to mathematically adjust for the profound non-stationarity and surveillance intensity biases introduced by the COVID-19 pandemic (fluctuating testing volumes and NPIs).

- Introducing our dual-stage solution: integrating DLNM and LSTM to forecast the Influenza Positivity Rate (rather than raw cases), structurally augmented with masking and testing volume covariates.

We believe this robust literature review now unequivocally establishes the necessity and novelty of our study.

> *(3) In connection with the above point, why the authors required the prospective cohort to assess a historical outcome should be highlighted clearly, which is one of the selling points of the study.*

You astutely highlight one of the core methodological strengths of our study design, and we appreciate the opportunity to emphasize this "selling point."

As noted in our response to Reviewer #1 regarding timeline splits, the phrase "prospective cohort to assess historical outcomes" reflects the administrative timeline of our study, but mathematically, the data stream is a continuous, longitudinal ecological surveillance.

The critical selling point here, which we have now explicitly highlighted in the revised Methods section, is that our "historical" data (2018–2020) was not collected via traditional, unstructured retrospective chart reviews. Instead, it was derived from an already operational, highly standardized public health sentinel surveillance system. Because this system utilized identical clinical case definitions, swabbing protocols, and RT-PCR diagnostic assays continuously from 2018 through 2023, the historical data inherently possesses the high fidelity, standardized quality, and lack of recall bias typically reserved for strict prospective cohorts. We have modified the Ethics statement subsection to explicitly underscore this epidemiological advantage.

"A major methodological strength of this study lies in its robust, uninterrupted longitudinal data collection framework spanning January 1, 2018, to December 31, 2023. While the analytical timeline encompasses a "retrospective" phase (January 1, 2018 – October 13, 2020) prior to formal ethical approval, and a "prospective" phase thereafter, we emphasize that this distinction represents a purely administrative demarcation regarding the timing of ethical approval. It does not reflect any shift in demographic cohorts, sentinel hospital locations, or data collection methodologies. Importantly, the historical data (2018–2020) were not subjected to the recall biases or misclassification risks typical of traditional retrospective chart reviews. Rather, they were systematically extracted from a continuously operating, highly standardized public health sentinel surveillance network. From the inception of data collection through the end of 2023, the local CDC maintained absolute uniformity in clinical influenza-like illness (ILI) definitions, nasopharyngeal swabbing procedures, and real-time reverse transcription polymerase chain reaction (RT-PCR) diagnostic assays. Consequently, the pre-2020 data possess the high-fidelity characteristics of a strict prospective cohort, ensuring unparalleled longitudinal consistency and mitigating temporal measurement bias across the entire pre-pandemic, pandemic, and postrestriction timeline." (Methods, page 8-9)

> *(4) The authors retrieved the daily data on the cases, which is usually small in number for most of the time, can often be driven by the importation (by population mobility with risk of infections) for particularly in a small location like a city.*

We deeply appreciate this incisive epidemiological observation. We fully agree that in a municipal-scale study, relying solely on daily absolute case counts presents significant mathematical and epidemiological vulnerabilities: absolute numbers can be small, highly stochastic, and susceptible to sudden spikes driven by imported cases rather than indigenous climate-driven transmission.

Driven directly by your comment, we have implemented two fundamental, structural upgrades to our study design:

- Shift to Positivity Rates: As detailed in our previous responses, we have entirely abandoned daily absolute case counts. Our DLNM and LSTM models now strictly utilize daily influenza positivity rates (positive cases ÷ total daily tested samples) for influenza A and B as the primary outcome. Positivity rates inherently smooth out the stochastic noise of small daily counts and provide a robust, normalized metric of true transmission intensity.

- Explicit Modeling of Importation Risk: To directly address the risk of importation via population mobility, we have integrated a novel covariate into our LSTM network: the daily proportion of non-local population/residents tested (labeled as outsiders_proportion in our SHAP analysis). By explicitly feeding this mobility proxy into the deep learning algorithm, the model is now mathematically equipped to contextualize and partial out the influence of imported infections when forecasting local transmission trends.

> *(5) The authors have not considered this extrinsic factor in the account and not even discussed it.*

We apologize for previously neglecting this critical extrinsic factor. As outlined in our response to Recommendation 4, we have now explicitly operationalized this extrinsic factor by incorporating daily non-local population proportion (outsiders_proportion) as a dynamic input feature in our revised LSTM architecture.

Furthermore, to empirically assess the magnitude of this importation risk, we conducted a retrospective analysis of the demographic data spanning our six-year study period. Our descriptive statistics reveal that days where the non-local population accounted for >50% of the daily tested cohort represented less than 1% of the total study days.

This empirical finding allows us to draw two important conclusions: First, while importation undoubtedly occurs (and is now accounted for by our LSTM covariate), indigenous transmission remains the overwhelmingly dominant driver of the observed epidemic curves in Putian. Second, massive importation shocks are rare enough that they do not systematically skew the overarching climate-disease associations identified by our models. We have thoroughly integrated both the methodological adjustment and this empirical discussion into the revised Methods and Discussion sections.

We invite you to review the corresponding manuscript clarifications.

“The study spanned January 1, 2018, to December 31, 2023, integrating four categories of data sources: (i) ILI and laboratory-confirmed cases from seven influenza sentinel hospitals across Putian’s urban and rural areas, ensuring representative coverage of diverse healthcare-seeking populations; (ii) daily meteorological data; (iii) COVID-19 associated public health intervention indicators (mask-wearing stringency indices) recorded from January 2020 onwards; and (iv) demographic mobility indicators (non-local population proportion) obtained from ILI consultation records.” (Methods, page 11)

“Beyond meteorological factors, our DLNM-LSTM framework systematically incorporated influenza positivity rates as the primary outcome variable, weekly detection volumes, mask-wearing stringency indices (indicator of NPIs during the COVID-19 pandemic), day of the week (DOW, distinguishing weekdays from weekends), and non-local population proportion (defined as the ratio of the number of individuals whose reported residential district at the time of testing lies outside Putian city to the total number of tests) as covariates within both DLNM and LSTM modeling pipelines. This comprehensive covariate framework ensures that observed meteorological associations are estimated after controlling for surveillance

intensity, public health intervention status, behavioral healthcare-seeking cycles, and population mobility patterns." (Methods, page 12)

Multivariate DLNMs efficiently expose transparent, lag-resolved main-effect surfaces for each meteorological variable, but cannot accommodate high-dimensional interactions among meteorological, autoregressive, and socio-behavioral factors without parameter inflation and severe multicollinearity. The LSTM stage was therefore not intended to replace DLNM inference, but to complement it by learning joint non-linear structure across concurrent covariates.

Crucially, this LSTM stage explicitly incorporates weekly detection volumes, maskwearing stringency indices, non-local population proportion, and DOW effects alongside meteorological inputs, and uses influenza positivity rates rather than absolute case counts as the modeling endpoint. Together, these design choices are intended to mitigate, rather than fully eliminate, the surveillance-related biases that can distort count-based forecasting during periods of fluctuating testing intensity." (Discussion, page 39-40)

> *(6) Further, how could such a small number of cases (which can be sporadic) define the epidemic onset and its uncertainty?*

You are absolutely correct: defining an epidemic onset using a small, sporadic number of absolute daily cases introduces severe statistical uncertainty and false-positive onset triggers. This specific methodological vulnerability was a primary catalyst for our decision to fundamentally pivot our analytical framework from absolute cases to Influenza Positivity Rates.

Unlike absolute counts, where a jump from 1 to 5 sporadic cases might artificially trigger an "onset" definition, positivity rates provide a continuous, normalized epidemiological signal. By assessing the proportion of positive tests against the total testing denominator, positivity rates mathematically stabilize the variance caused by sporadic daily testing. Consequently, an upward trajectory in positivity rates provides a highly reliable, low uncertainty signal of true epidemic onset and acceleration. The exceptional validation metrics of our revised LSTM model (e.g., MAE of 0.009 for Influenza A positivity rate) demonstrate that utilizing this normalized metric virtually eliminates the noise and uncertainty associated with sporadic small-number counts.

> *(7) Figure 3 presents the time series of the cases. I wonder whether the data for these factors and outcomes are daily or aggregated by week/month? I suggest representing it in 9x1 format with a single x-axis to compare, instead of 3x3 format. Authors can refer similar plot in https://doi.org/10.1371/journal.pcbi.1012311 in Figure 1.*

We are extremely grateful for this specific and highly constructive visualization suggestion. To answer your query: the data plotted for both the meteorological factors and the influenza outcomes are indeed daily observations.

We fully agree that the original 3x3 format severely compromised the readability of this daily data. Following your excellent advice and referencing the suggested literature, we have entirely redesigned Figure 3 into an 8x1 vertically stacked format with a single shared continuous x-axis. This structural upgrade has completely transformed the figure, eliminating the horizontal compression and elegantly exposing the fine-grained, daily temporal alignments between climatic extremes and viral surges. The newly rendered Figure 3 is now much more intuitive and analytically valuable. Due to space constraints and given that the revised Figure 3 has been included in our response to a similar comment from Reviewer #1, we will not reproduce the figure in this response. We invite you to review the revised manuscript or refer to our response to Reviewer #1's recommendation 1 for Figure 3.

> *(8) "Additionally, we plotted the loss function curves for the network on the training and validation sets to monitor LSTM convergence and the risk of overfitting." The authors validated the model with predefined training and validation sets. I suggest providing more details on the techniques and the length of the sets. How are these considerations safe for the assumptions and limitations of the models?*

We sincerely thank you for this crucial request for methodological transparency. You are absolutely correct that the techniques used for data partitioning are fundamental to the safety and validity of time-series modeling assumptions. To strictly respect the temporal dependencies of LSTM networks and avoid future-to-past data leakage, we avoided standard random train/test splitting. Instead, we implemented strict Chronological Out-of-Time (OOT) Three-Tier Validation Architecture.

We have extensively expanded the Methods and Discussion sections to detail this partitioning scheme, our rationale, and its associated limitations:

- The Three-Tier Validation Architecture and Length of Sets:

Tier 1 — Training and Internal Cross-Validation (2018–2022, ~88%): Used for initial model fitting. Within this phase, 10% of the samples were held out during Bayesian optimization as an internal validation fold strictly for monitoring convergence, controlling overfitting (via early stopping), and guiding the hyperparameter search. This internal fold never contributed to the final reported performance metrics.

Tier 2 — Internal Hold-out Test Set (2023, ~12%): A continuous 365-day block reserved as a strictly chronological hold-out, used solely for final performance evaluation. No information from this period influenced training or hyperparameter tuning.

Tier 3 — External Independent Validation (Sanming 2023): The full surveillance time series from a geographically distinct subtropical city, providing the strongest evidence of cross-location generalizability.

- Justification of the 88% / 12% Partition Ratio:

This specific ratio was deliberately chosen to balance two competing epidemiological and computational considerations:

Sufficient Training Memory: The model required a multi-year training continuum (2018–2022) to autonomously learn the structural breaks and complex non-stationarities introduced by the COVID-19 pandemic and strict NPIs.

Epidemiological Gold Standard for Testing: Allocating exactly one year (2023) for testing is the epidemiological gold standard for seasonal infectious diseases. A full 365-day cycle ensures that model performance is evaluated across all seasonal phases (spring peaks, summer lulls, winter rebounds) rather than a biased, partial-year fragment.

- Methodological Safeguards Against Information Leakage (Safe Assumptions):

To ensure these partitions were safe for the model's assumptions, we implemented strict safeguards:

No Temporal Leakage: Tier 2 strictly follows Tier 1 in calendar time, preserving the sequential integrity assumed by LSTM architectures.

No Distributional Leakage: All feature scaling and normalization parameters (means, standard deviations, min-max ranges) were derived exclusively from Tier 1 (Training) and applied unchanged to Tiers 2 and 3.

- Explicit Acknowledgement of Limitations:

We honestly acknowledge that while fixed chronological partitioning is the methodological standard for LSTM forecasting, a single fixed split point (Dec 31, 2022) fundamentally tests only one structural break. It does not exhaustively probe all potential future non-stationarities. Alternative strategies, such as expanding-window or rolling origin cross-validation, could offer additional robustness characterizations. We have integrated this crucial point into the Limitations section of the Discussion.

We invite you to review the corresponding manuscript clarifications.

“To respect the temporal dependence inherent in LSTM architectures and avoid data leakage, we adopted a strict chronological out-of-time (OOT) validation strategy. Timeseries data from January 1, 2018, to December 31, 2022 (88.26% of the Putian dataset) were used for model training, with 10% reserved during Bayesian optimization as an internal validation subset for convergence monitoring and hyperparameter tuning only. Data from January 1, 2023, to December 31, 2023 (11.74%) were held out as a chronologically internal validation set for final performance evaluation, covering a complete annual cycle. External validation was conducted using concurrent 2023 data from Sanming city to assess spatial generalizability. To prevent distributional leakage, all normalization parameters were derived exclusively from the training set and consistently applied to the validation sets, with predictions subsequently transformed back to the original scale.” (Methods, page 18)

“Secondly, the interpretation of our framework’s predictive performance during the 2023 validation period requires careful epidemiological and methodological contextualization. The year of 2023 represented an anomalous, post-restriction “rebound” period characterized by rapid NPI relaxation and the release of accumulated population-level immunity debt, resulting in an atypical influenza surge that exceeded pre-pandemic peaks. The framework’s high accuracy across this period should therefore be interpreted as evidence of algorithmic agility and adaptive capacity during a highly volatile transitional phase, rather than as definitive proof of long-term predictive validity under a stabilized post-2024 epidemiological regime. Methodologically, while our strict chronological OOT data partitioning prevented temporal information leakage, a critical requirement for LSTM integrity, the reiance on a single, fixed chronological split point (December 31, 2022) intrinsically limits our evaluation to one specific structural break. This fixed-split approach may not exhaustively probe the DLNM-LSTM framework’s resilience against all forms of future epidemiological non-stationarity. Consequently, naive extrapolation of the reported 2023 performance metrics to future surveillance years should be avoided absent prospective recalibration. Future studies should consider employing expanding-window or rolling-origin cross-validation frameworks to provide a more continuous characterization of algorithmic robustness. Continuous integration of accumulating 2024 and 2025 data, combined with adaptive learning architectures capable of detecting regime shifts in real time, will be essential before any operational deployment of this, or similar forecasting frameworks, for routine public health surveillance.” (Discussion, page 44-45)

> *(9) The authors considered the DLNM analysis without considering the potential interactions among meteorological factors, which can't be avoided in real-world environmental contexts. I would suggest constructing such models by incorporating more reasonable interaction terms to reflect the complex relationships between these variables. Although the impact of COVID-19 was considered on the outcome of influenza directly.*

We sincerely thank you for raising this vital conceptual point. We completely agree that meteorological factors exhibit physically real interactions under real-world environmental conditions (e.g., the synergistic effect of extreme heat and high humidity on viral viability and

aerosol dynamics). We welcome the opportunity to clarify how our analytical framework structurally addresses this precise complexity without compromising mathematical stability.

Rather than forcing interaction terms into a single statistical model, we designed our dual-stage architecture specifically to create a functional division of labor between the DLNM and LSTM frameworks:

- Methodological Constraints on Explicit DLNM Interaction Terms: While conceptually appealing, explicitly incorporating pairwise interaction terms across eight meteorological variables within the DLNM stage would generate 28 two-way interaction cross-bases, each with its own non-linear and lag-distributed spline structure. This parameter explosion leads to the "curse of dimensionality," causing: (a) severe multicollinearity given the strong baseline correlations among weather variables; (b) profound instability of the cross-basis estimates and inflated standard errors; (c) a massive risk of overfitting; and (d) the complete loss of visual interpretability, which is the principal value proposition of DLNM. Consequently, as is standard practice in environmental epidemiology (Gasparrini et al., 2010), we restricted our DLNM stage to isolating interpretable, lag-distributed main-effect exposure–response surfaces.

- The LSTM Stage as the Engine for Complex Interactions: This is precisely where the deep learning architecture provides its unique methodological value. The LSTM network does not require analysts to manually pre-specify rigid interaction terms. Instead, its multi-layer, non-linear gating architecture is intrinsically capable of autonomously extracting and representing arbitrary, high-dimensional interactions among all input variables simultaneously.

Therefore, complex meteorological interactions are absolutely not ignored in our study; rather, they are absorbed into and resolved by the LSTM stage to maximize predictive accuracy, while the DLNM stage provides the lag-resolved, interpretable backbone for individual main effects. Our multi-stage variable integration ensures that the limitations of traditional statistical models do not artificially bottleneck the deep learning framework's capacity to synthesize real-world complexities.

We have now added explicit paragraphs to both the Methods and Discussion sections articulating this functional division of labor, ensuring maximum methodological transparency regarding how meteorological interactions are accommodated within our framework.

We invite you to review the corresponding manuscript clarifications.

"In the first stage, DLNMs characterize subtype-specific, non-linear, and lag-distributed associations between meteorological variables and influenza A and B positivity. In the second stage, a Bayesian-optimized LSTM network integrating meteorological, autoregressive, and socio-behavioral covariates is used for short-horizon forecasting, benchmarked against a covariate-matched multivariate ARIMA model and evaluated in an independent subtropical city (Sanming) as a preliminary test of model portability. Influenza positivity rate is used as the primary modeling endpoint to mitigate testing-related surveillance bias. Our aim is to provide an interpretable, climate-informed forecasting approach for subtropical influenza that can serve as a methodological foundation for future operational surveillance developments." (Introduction, page 7-8)

"DLNM was first constructed to screen meteorological factors and other covariates with substantial influence on influenza seasonality. Subsequently, an LSTM neural network was developed within the same covariate system. The integrated DLNM–LSTM framework was employed as methodologically complementary rather than redundant components, leveraging the distinctive strengths of each approach to address different analytical

objectives within a unified investigation. Specifically, DLNM models characterize the nonlinear exposure–lag–response relationships between meteorological factors and influenza risk, providing biologically interpretable insights into the temporal structure of weather influenza associations and identifying meteorological factors with statistically and clinically significant effects on influenza dynamics." (Methods, page 11)

"Building on the significant non-linear and lagged effects through DLNM analysis of real world environmental exposures, this study further constructed multi-factor influenza A and B prediction LSTM networks. Leveraging a recurrent architecture with non-linear gating mechanisms, these networks are able to automatically capture and represent complex, high dimensional interactions among meteorological variables without the need for manual prespecification. This functional division of labor between the DLNM and LSTM models enhances predictive performance while preserving the interpretability and inferential stability established in the DLNM stage." (Methods, page 18)

"A central methodological feature of our framework is the deliberate division of labor between the DLNM and LSTM components. Multivariate DLNMs efficiently expose transparent, lag-resolved main-effect surfaces for each meteorological variable, but cannot accommodate high-dimensional interactions among meteorological, autoregressive, and socio-behavioral factors without parameter inflation and severe multicollinearity. The LSTM stage was therefore not intended to replace DLNM inference, but to complement it by learning joint non-linear structure across concurrent covariates. Furthermore, extensive environmental inputs inevitably introduce severe collinearity, such as the strongly correlated solar radiation and UV index. While traditional multivariate models are highly vulnerable to such overlapping variances, the recurrent, weighted representation learned by the LSTM is comparatively tolerant of such redundancy, allowing broader covariate integration than in previous efforts (Zhu et al. 2022). Crucially, this LSTM stage explicitly incorporates weekly detection volumes, mask-wearing stringency indices, non-local population proportion, and DOW effects alongside meteorological inputs, and uses influenza positivity rates rather than absolute case counts as the modeling endpoint. Together, these design choices are intended to mitigate, rather than fully eliminate, the surveillance-related biases that can distort count-based forecasting during periods of fluctuating testing intensity." (Discussion, page 39-40)

> *(10) In context with the above points, although COVID-19-related variables are included, important confounding factors such as population mobility, vaccination coverage, and school calendar (e.g., school openings/closings) are not adequately considered. Additionally, there is a potential risk of overfitting due to an imbalanced data split-too much data is allocated to the training set, while the validation set is relatively small on the other hand.*

We sincerely appreciate your comprehensive evaluation regarding confounding control and the risk of overfitting. These are highly pertinent methodological concerns, and we have implemented multiple refinements and empirical justifications to address each of them systematically.

- Comprehensive Control of Confounding Factors

To address the omitted confounders you rightfully identified, we have substantially expanded our covariate framework in the revised models:

Population Mobility: We introduced the proportion of the migrant/non-local population as a new quantitative covariate (computed as the ratio of tested individuals reporting non-Putian residential addresses). This explicitly captures the extrinsic transmission pressure and viral importation risk exerted by mobile populations.

Social/School Routines: To capture cyclical social contact patterns and surveillance reporting

dynamics, we incorporated the Day-of-the-Week (DOW) indicator (distinguishing weekdays from weekends).

Vaccination Coverage & School Calendar (Limitations): We honestly acknowledge that highly granular, municipal-level daily vaccination registry data and official macro-school holiday timelines were unavailable for integration into our daily time-series framework. However, for essential epidemiological context, the overall influenza vaccination coverage in mainland China during the 2018–2023 study window is historically estimated at a mere 2% to 3% of the general population, substantially lower than the 40–60% coverage typically observed in high-income temperate countries. Given this exceedingly low baseline, the population-level confounding contribution of vaccination on our predictive accuracy is expected to be minimal in absolute magnitude. We have now explicitly addressed this specific regional epidemiological context in the Discussion section.

- Justification of the Data Split Ratio (88% vs. 12%)

Regarding the perceived imbalance in the data partition, the chronological 88% / 12% split was not arbitrary; it was designed to balance two competing epidemiological necessities:

Preserving Training Memory: The 88% training block (2018–2022) was strictly required for the LSTM to autonomously learn the multi-year seasonal cycles, the structural breaks induced by COVID-19 NPIs, and the suppressed-regime dynamics.

The Epidemiological Gold Standard: The 12% testing block equates exactly to the 2023 calendar year (365 days). In seasonal infectious disease forecasting, evaluating performance across a complete, unbroken annual cycle is the epidemiological gold standard, ensuring the model is tested across all phases (spring peaks, summer lulls, winter rebounds) rather than a biased, partial-year fragment.

- Safeguards Against Overfitting & Empirical Proof (Sensitivity Analysis)

To definitively mitigate and disprove the risk of overfitting, we implemented a Chronological Three-Tier Validation Architecture:

During the training phase, we randomly partitioned a 10% internal validation subset exclusively for hyperparameter tuning (via Bayesian Hyperopt) and implementing early stopping to terminate training the moment validation loss plateaued.

We utilized the full 2023 surveillance data from an entirely distinct city (Sanming) as an external independent validation set. The fact that our model generalized excellently to Sanming is the strongest empirical proof against localized overfitting.

Finally, to further substantiate model robustness, we conducted a controlled covariateablation sensitivity analysis (new Table 2). If a deep learning model is severely overfitted (i.e., memorizing noise), removing covariates often yields chaotic or random performance changes. However, when we explicitly removed the “mask-wearing stringency indices” or the “weekly detection volumes”, our model’s performance degraded in a predictable, biologically plausible manner (e.g., MAE increased by 33.3% to 100% across subtypes).

This structurally proves that our LSTM architecture is not achieving artificially robust performance through overfitting, but exhibits appropriate sensitivity to the exact epidemiological features driving true viral transmission.

We have extensively documented these justifications, safeguards, and limitations in the revised Methods, Results, and Discussion sections.

We invite you to review the newly added limitation statement paragraph.

"Thirdly, beyond the surveillance coverage limitation noted above, the aggregate-level nature of the available data further constrained our ability to adjust for individual-level confounders, including personal vaccination status, detailed comorbidities, healthcare seeking behavior, and socioeconomic status. This limitation was particularly exacerbated by the profound epidemiological disruptions during the COVID-19 pandemic, where raw numbers of confirmed cases became heavily confounded by surveillance intensity (i.e., fluctuating testing volumes) rather than solely reflecting underlying viral transmission. Our adoption of influenza positivity rates as the primary modeling endpoint and incorporation of weekly detection volumes, face-covering stringency, and non-local population proportion as dynamic covariates, rigorously mitigated these aggregate-level biases and linked our methodological design directly to the forecasting outcomes. As unequivocally demonstrated by our covariate-ablation sensitivity analyses (Table 2), failing to account for mask mandates and testing volumes leads to severe, mathematically predictable deviations in absolute forecasting accuracy. Furthermore, while daily case counts in a single city can occasionally be small, sporadic, and driven by external importations, our incorporation of non-local population proportion effectively adjusted for these localized importation risks. Consequently, although our findings characterize population-level associations between meteorological factors and influenza activity, they should not be interpreted as evidence of micro-level causal mechanisms at the individual patient level. Ultimately, the DLNMLSTM framework's robust performance across the non-stationary transition out of NPI policies highlights the absolute necessity of integrating behavioral and virological baseline metrics into future climate-driven predictive surveillance systems." (Discussion, page 45-46)

> *(11) The authors should justify why the baseline model selection was made by comparing the LSTM model only with ARIMA? How the outcomes could be sensitive to other commonly used machine learning methods, such as Random Forest or XGBoost, etc, as a benchmark for their performance.*

We are deeply grateful for this methodologically incisive suggestion, which prompted us to substantially strengthen the manuscript's benchmarking framework. Recognising the methodological importance of this recommendation, our team initiated the construction of an eXtreme Gradient Boosting (XGBoost) benchmark model in parallel with the Round 1 revision submission, and has now completed its full validation and interpretive analysis. XGBoost, one of the leading non-deep-learning machine-learning frameworks for structured predictive tasks, was implemented using the identical covariate matrix employed in the LSTM. The complete methodology, results, and interpretive analysis have been incorporated as Supplementary Additional File 6, with corresponding updates to the Methods (Study Design and Model Construction), Results, and Discussion sections of the main manuscript.

Methodological design. The XGBoost model was implemented in R (xgboost package, version 3.2.1.1) and received a covariate set strictly matched to the LSTM: eight meteorological variables, weekly testing volumes, the mask-wearing stringency index, a day-of-week indicator, and the proportion of non-local (migrant) population. Critically, no additional lag features, sliding-window statistics, or autoregressive terms were introduced, ensuring that XGBoost's information set was identical to the contemporaneous covariate stream consumed by the LSTM at each time step. This design deliberately isolates the predictive contribution of architectural inductive bias, the LSTM's intrinsic sequential memory versus XGBoost's memory-free tree-partitioning geometry, from any confounding due to hand-engineered temporal features. The training (2018 – 2022) and validation (2023) partitions and the evaluation metrics (MAE, RMSE, MAPE, SMAPE) followed the LSTM and ARIMA protocols precisely.

Empirical findings.

A complete performance matrix including MAE, RMSE, MAPE, and SMAPE for all three models is presented in Supplementary Table S3 (Additional File 6), which confirms the same rank-ordering (LSTM ≪ ARIMA ≈ XGBoost) across all four error metrics.

Despite receiving identical inputs, XGBoost yielded errors approximately 15-fold higher than the LSTM for Influenza A and 35-fold higher for Influenza B, with errors broadly comparable in magnitude to the ARIMA baseline. Visual inspection of the 2023 forecasting trajectories (Supplementary Figure S6) revealed that XGBoost systematically under-predicted the explosive post-NPI Influenza A rebound in late 2023 and generated largely flat forecasts for Influenza B that failed to reproduce the mid-year peak.

Interpretation. These findings are mechanistically informative and reinforce, rather than merely confirm, the manuscript's central methodological claim. Four architectural properties of XGBoost account for the observed performance gap:

Insensitivity to temporal ordering. Decision-tree splits operate on feature-dimensional thresholds and cannot structurally encode "the present depends on the past" in the manner natively accommodated by recurrent architectures.

Absence of gated recurrence. XGBoost lacks any mechanism analogous to the LSTM's forget-input-output gating for propagating, filtering, and dynamically re-weighting historical states, and therefore cannot represent long-range non-linear temporal dependencies.

Violated independence assumption. Tree ensembles implicitly assume independent and identically distributed (i.i.d.) observations, an assumption fundamentally at odds with the strong serial autocorrelation of epidemiological time series.

Non-comparable modelling paradigms. Both ARIMA and LSTM are explicit sequential models within a common paradigm, whereas XGBoost belongs to a fundamentally non-sequential family. The LSTM–ARIMA contrast therefore isolates the specific contribution of deep sequential learning within a paradigmatically comparable framework, while the LSTM–XGBoost contrast demonstrates that even a state-of-the-art memory-free non-linear learner cannot substitute for a genuinely sequential architecture under epidemiologically non-stationary conditions.

The consistent superiority of the LSTM over both a linear statistical benchmark (ARIMA) and a non-linear memory-free machine-learning benchmark (XGBoost), all receiving identical inputs and evaluated on the identical validation window, isolates the recurrent gated architecture itself as the source of the predictive gain. We are indebted to the Reviewer for prompting this three-way benchmark analysis, which has materially strengthened the methodological rigour and interpretive depth of the manuscript.

This additional benchmark analysis, though completed subsequent to the Round 1 submission, has now been fully integrated into the Round 2 revision at the appropriate positions across the Abstract, Importance Statement, Methods, Results, Discussion, and Conclusion, together with the new Additional File 6. We are grateful to you for prompting this three-way benchmark, which we believe has materially strengthened the methodological rigour and interpretive depth of the manuscript.

We invite you to review the newly added Discussion paragraphs.

"The optimized LSTM algorithm demonstrated strong predictive performance, with loss curves for both subtype-specific networks exhibiting favourable convergence, consistent with prior work (Du et al. 2023). To rigorously isolate the predictive contribution attributable to the LSTM's recurrent architecture, we benchmarked it against two covariate-matched baselines representing distinct methodological paradigms: a multivariate ARIMA model and

an XGBoost gradient-boosting ensemble. This design controls simultaneously for linearity (ARIMA → LSTM contrast) and for non-linearity without recurrent memory (XGBoost → LSTM contrast), allowing us to attribute observed performance gains to specific architectural inductive biases rather than to informational asymmetry or model non-linearity in general. The LSTM substantially outperformed both baselines on the 2023 validation window. For influenza A, it achieved an MAE of 0.009, compared with 0.136 for ARIMA and 0.138 for XGBoost, approximately 15-fold reductions. For influenza B, the LSTM yielded an MAE of 0.002, versus 0.049 for ARIMA and 0.070 for XGBoost, 25- to 35-fold reductions. Critically, ARIMA and XGBoost produced errors of comparable magnitude despite their disparate assumptions regarding linearity, and both systematically under-predicted the explosive 2023 post-NPI rebound for influenza A while generating flat or spurious trajectories for influenza B (Supplementary Figures S5–S6). That parallel failure indicates the LSTM's advantage under pandemic-era non-stationarity derives not from non-linearity per se, but from four architectural properties intrinsic to recurrent gated networks yet absent in tree ensembles: (i) threshold-based, sequence-insensitive splits fail to encode present–past dynamics; (ii) XGBoost lacks forget–input–output gates to propagate and re-weight historical states across time lags; (iii) gradient-boosted trees assume near-independence, contradicting the pronounced temporal autocorrelation in epidemiological time series; (iv) ARIMA and LSTM are sequential models with different functional forms, whereas XGBoost is non-sequential. The joint failure of ARIMA and XGBoost, despite differing non-linear treatment, isolates recurrent sequential memory as the critical architectural feature for forecasting under non-stationarity. We emphasize that this interpretation applies specifically to the present non-stationary influenza forecasting task and does not constitute a general dismissal of gradient-boosted ensembles, which retain state-of-the-art performance across many structured prediction domains. Rather, it highlights that for surveillance time series exhibiting pronounced temporal dependencies and abrupt regime shifts, such as the 2023 post-NPI rebound, explicit sequential memory becomes functionally essential. This mechanistic reading, together with the LSTM's comparative edge over previously reported ARIMA-based (Li et al. 2024) and LSTM-based influenza prediction models (Zhu et al. 2022), positions our framework as a substantive methodological advance in predictive modeling for climate-sensitive diseases." (Discussion, page 40-42)

"Fourthly, our benchmarking strategy was deliberately structured as a paradigmatically layered three-way comparison (linear-sequential ARIMA, non-linear non-sequential XGBoost, non-linear sequential LSTM) rather than an exhaustive algorithmic survey. This design prioritized methodological clarity, isolating the contribution of recurrent sequential inductive bias, over horizontal coverage. Nonetheless, our evaluation does not extend to Transformer-based attention architectures, nor to hybrid ensemble strategies (e.g., LSTM–XGBoost stacking or multi-model Bayesian model averaging) that may offer complementary strengths. Systematic benchmarking against these emerging architectures, alongside prospective recalibration on additional subtropical surveillance streams and operational stress-testing under real-time data latency, will be essential next steps as the framework evolves toward operational deployment."(Discussion, page 46)

> *(12) I was expecting the statement of generalizability in terms of locations with the link to the results of the study. I suggest including such statements in the text along with limitations.*

We sincerely thank you for this highly constructive suggestion. We fully agree that an explicit, results-linked generalizability statement is crucial for appropriately interpreting and safely deploying our forecasting framework. To address this, we have substantially expanded the Limitations subsection within the Discussion to articulate a rigorous, three tier generalizability framework directly linked to our study findings:

- Empirically Validated Generalization: The successful external validation of our framework in Sanming, a geographically distinct, inland prefecture-level city, constitutes direct empirical evidence of cross-location generalizability within the subtropical south eastern Chinese context. As detailed in our results, the framework retained high operational forecasting accuracy (e.g., Influenza A MAE of 0.015, SMAPE of 0.610) without any retraining on Sanming's local data, proving that the meteorological exposure–response patterns and LSTM architecture are transferable across similar subtropical climates under analogous public health intervention frameworks.

- Plausible Inferential Generalization: Beyond the directly validated Putian–Sanming pair, the framework's architecture is plausibly generalizable to other subtropical cities across adjacent regions (e.g., Guangdong, eastern Guangxi) that share comparable monsoon climates, year-round multi-peak influenza circulation patterns, and similar sentinel surveillance infrastructures.

- Boundaries of Transferability (Non-generalizability): We explicitly state the boundaries where our specific model parameters should not be directly extrapolated without rigorous local recalibration. These include: (a) regions located in subtropical latitudes but characterized by fundamentally distinct climate regimes (e.g., monsoonal systems with characteristics that are markedly incomparable), where the temperature–humidity coupling structures and seasonality with multiple winter peaks differ qualitatively from those in the study setting; and (b) regions with substantially different public-health intervention landscapes, such as settings with high population-level vaccination coverage or school-closure-based mitigation strategies.

By explicitly defining these boundaries in the revised Discussion, we ensure transparent communication of the model's appropriate application scope and methodological limitations.

We invite you to review the newly added limitation statement paragraph.

"Furthermore, the Bayesian-optimized LSTM architecture was directly applied, without re-training or hyperparameter adjustment, to both the Putian internal test set and the Sanming external validation set. This unified modeling framework ensures that performance differences between the two evaluation contexts primarily reflect geographic transferability rather than model re-optimization." (Methods, page 21)

Finally, it is imperative to explicitly delineate the generalizability of our forecasting framework within geographic and epidemiological contexts. The successful external validation in Sanming provides direct empirical evidence that our LSTM architecture and the identified meteorological thresholds are robustly transferable across the subtropical southeastern Chinese context, sharing comparable monsoon climates, year-round influenza circulation, and standardized public health intervention frameworks. However, we explicitly define the boundaries of this transferability. The specific meteorological coefficients, lag structure, and predictive parameters derived in this study should not be indiscriminately extrapolated to other regions, even within subtropical latitudes, where climatic regimes (e.g., monsoon regimes with markedly non-comparable characteristics) or public health contexts (e.g., higher baseline influenza vaccination coverage or distinct non-pharmaceutical intervention strategies) differ substantially. In such disparate settings, directly applying our pretrained model may yield substantial systematic biases. While the underlying modeling framework remains methodologically transferable, its parameterization requires rigorous local recalibration using region-specific surveillance data. Acknowledging these boundaries ensures that the framework can be deployed more safely and appropriately to realize targeted, climate-sensitive infectious disease forecasts." (Discussion, page 44-47)"

> *(13) The flow of the manuscript should be revised for general readers, for example*

> *author spent a lot of text on limitations in general without linking them to the outcomes and study design. The English language has a huge scope for improvement. I suggest paying attention to the presentation of the text for the English language use and continuity.*

We sincerely appreciate your candid feedback on the manuscript's readability. We recognize that in our previous draft, integrating diverse critiques from multiple rounds of peer review inadvertently resulted in disjointed narrative flows, particularly in the Methods and Limitations sections.

To address this, we have undertaken a massive structural overhaul and linguistic refinement of the entire manuscript:

- Logical Flow: The Introduction has been sharply refocused; the Methods section has been streamlined (e.g., moving lengthy PCR protocols to the Supplement); and the Results have been reorganized with clear subheadings.

- Contextualized Limitations: We completely rewrote the Discussion section. Rather than listing generic limitations, we have deeply anchored them to our specific study design and outcomes. For instance, we now explicitly discuss how our shift to predicting the "Positivity Rate" directly mitigates the "Surveillance Bias" limitation, and how the 2023 "rebound" limits steady-state generalizations.

- Language Polish: The manuscript has undergone comprehensive editing by a native English-speaking academic expert to ensure grammatical precision, sophisticated vocabulary, and seamless continuity.

We believe these revisions have dramatically elevated the clarity and scholarly tone of the text.

https://doi.org/10.7554/eLife.107767.2.sa0