

# **From Binary to Probability: A Machine Learning-Based Early Warning System for Seasonal Influenza in China (2014-2024)**

Jinzhao Cui, Liuyang Yang, Ting Zhang, Xuefeng Huang, Xiaochen Zhang, Yongtao Chi, Qiang Huang, Yu Yang, Zhongjie Li, Xiaoli Wang, Weizhong Yang

Submitted to: JMIR Medical Informatics  
on: March 08, 2025

**Disclaimer:** © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

## ***Table of Contents***

---

**Original Manuscript..... 5**



# From Binary to Probability: A Machine Learning-Based Early Warning System for Seasonal Influenza in China (2014–2024)

Jinzhao Cui<sup>1\*</sup> PhD; Liuyang Yang<sup>2\*</sup> PhD; Ting Zhang<sup>3\*</sup> PhD; Xuefeng Huang<sup>4</sup> MD; Xiaochen Zhang<sup>4</sup> MD; Yongtao Chi<sup>4</sup> MD; Qiang Huang<sup>4</sup> MD; Yu Yang<sup>4</sup> MD; Zhongjie Li<sup>3</sup> Prof Dr Med; Xiaoli Wang<sup>5</sup> PhD; Weizhong Yang<sup>6</sup> Prof Dr Med

<sup>1</sup> Public Health Emergency Management Innovation Center Beijing CN

<sup>2</sup> Fudan University, Shanghai, China Shanghai CN

<sup>3</sup> 1. School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

2. Public Health Emergency Management Innovation Center, Beijing, China Beijing CN

<sup>4</sup> School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China Beijing CN

<sup>5</sup> Beijing Center for Disease Prevention and Control Beijing CN

<sup>6</sup> Chinese Academy of Medical Sciences & Peking Union Medical College Beijing CN

\* these authors contributed equally

## Corresponding Author:

Zhongjie Li Prof Dr Med

1. School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China 2. Public Health Emergency Management Innovation Center, Beijing, China

No. 31, Beijijgesantiao street, Dongcheng District, Beijing

Beijing

CN

## Abstract

**Background:** Seasonal influenza remains a major global public health concern, causing substantial morbidity and mortality. Traditional early warning models rely on binary (0/1) classification methods, issuing alerts only when predefined thresholds are crossed. However, these models lack flexibility, often leading to false alarms or missed warnings, and fail to provide granular risk assessments for decision-making. To address these limitations, we propose a probability-based early warning system using machine learning, offering continuous risk estimation (0-1 variable) instead of rigid threshold-based alerts.

**Objective:** We want to build an infectious disease prediction model of seasonal influenza based on machine learning method to provide theoretical basis for early warning of infectious diseases

**Methods:** We developed a Dense ResNet machine learning model trained on influenza surveillance data from Northern and Southern China (2014–2024). The model generates early warnings 3, 5, and 7 days in advance, providing a probability-based risk assessment (0-1 continuous variable) instead of traditional binary (0/1) warnings. We evaluated model performance using AUC scores, accuracy, recall, and F1 scores, comparing it with support vector machines (SVM), random forests, XGBoost, and LSTM models.

**Results:** The Dense ResNet model demonstrated the best performance with 5-day lead warnings and a 50th percentile probability threshold, achieving AUC scores of 0.94 (Northern China) and 0.95 (Southern China). Compared to traditional models, probability-based warnings improved early detection, reduced false alarms, and allowed for tiered public health responses.

**Conclusions:** This study introduces a probability-based influenza warning system, offering key advantages over traditional binary models. By providing a continuous risk assessment, the system enables refined decision-making rather than rigid threshold-based warnings. The flexibility of probability-based warnings supports tiered response strategies, allowing health authorities to adjust interventions dynamically based on the predicted risk level. Additionally, integrating AI-assisted automation enhances efficiency—higher probabilities (>0.7) can trigger automatic medical alerts, while lower probabilities (0.4-0.6) can be used for internal monitoring without unnecessary public alarm. Compared to fixed-threshold methods (e.g., 40th percentile warnings), this approach provides earlier detection, better adaptation to epidemic trends, and reduced false positives. The model's ability to issue warnings 5 days in advance offers a critical window for medical resource allocation, vaccination

strategies, and public health interventions.

This study presents a novel probability-based machine learning model for influenza early warning, demonstrating superior accuracy, flexibility, and practical applicability. By replacing binary warnings with probability-driven risk assessments, this approach enhances influenza preparedness and supports automated AI-driven public health responses. Future research should integrate real-time surveillance data and transmission dynamic models to further improve early warning precision. Clinical Trial: null

(JMIR Preprints 08/03/2025:73631)

DOI: <https://doi.org/10.2196/preprints.73631>

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

**Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

**Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in a JMIR journal, I will be able to make my accepted manuscript PDF available to anyone.

No. Please do not make my accepted manuscript PDF available to anyone.

**Original Manuscript**



## Methods and Applications

# From Binary to Probability: A Machine Learning-Based Early Warning System for Seasonal Influenza in China (2014–2024)

*Jinzhao Cui<sup>1†</sup>, Liuyang Yang<sup>2†</sup>, Ting Zhang<sup>1†</sup>, Xuefeng Huang<sup>1</sup>, Xiaochen Zhang<sup>1</sup>, Yongtao Chi<sup>1</sup>, Qiang Huang<sup>1</sup>, Yu Yang<sup>1</sup>, Zhongjie Li<sup>1\*</sup>, Xiaoli Wang<sup>3\*</sup>, Weizhong Yang<sup>1</sup>*

<sup>1</sup> School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

<sup>2</sup> Fudan University, Shanghai, China

<sup>3</sup> Beijing Center for Disease Prevention and Control, Beijing, China

<sup>†</sup>These authors contributed equality.

<sup>\*</sup>To whom correspondence should be addressed to: [lizhongjie@sph.pumc.edu.cn](mailto:lizhongjie@sph.pumc.edu.cn) (ZhongJie Li)

## Abstract

**Background:** Seasonal influenza remains a major global public health concern, causing substantial morbidity and mortality. Traditional early warning models rely on binary (0/1) classification methods, issuing alerts only when predefined thresholds are crossed. However, these models lack flexibility, often leading to false alarms or missed warnings, and fail to provide granular risk assessments for decision-making. To address these limitations, we propose a probability-based early warning system using machine learning, offering continuous risk estimation (0-1 variable) instead of rigid threshold-based alerts.

**Methods:** We developed a Dense ResNet machine learning model trained on influenza surveillance data from Northern and Southern China (2014–2024). The model generates early warnings 3, 5, and 7 days in advance, providing a probability-based risk assessment (0-1 continuous variable) instead of traditional binary (0/1) warnings. We evaluated model performance using AUC scores, accuracy, recall, and F1 scores, comparing it with support vector machines (SVM), random forests, XGBoost, and LSTM models.

**Results:** The Dense ResNet model demonstrated the best performance with 5-day lead warnings and a 50th percentile probability threshold, achieving AUC scores of 0.94 (Northern China) and 0.95 (Southern China). Compared to traditional models, probability-based warnings improved early detection, reduced false alarms, and allowed for tiered public health responses.

**Discussion:** This study introduces a probability-based influenza warning system, offering key advantages over traditional binary models. By providing a continuous risk assessment, the system enables refined decision-making rather than rigid threshold-based warnings. The flexibility of probability-based warnings supports tiered response strategies, allowing health authorities to adjust interventions dynamically based on the predicted risk level. Additionally, integrating AI-assisted automation enhances efficiency—higher probabilities ( $\geq 0.7$ ) can trigger automatic medical alerts, while lower probabilities (0.4-0.6) can be used for internal monitoring without unnecessary public alarm. Compared to fixed-threshold methods (e.g., 40th percentile warnings), this approach provides earlier detection, better adaptation to epidemic trends, and reduced false positives. The model's ability to issue warnings 5 days in advance offers a critical window for medical resource allocation, vaccination strategies, and public health interventions.

**Conclusion:** This study presents a novel probability-based machine learning model for influenza early warning, demonstrating superior accuracy, flexibility, and practical applicability. By replacing binary warnings with probability-driven risk assessments, this approach enhances influenza preparedness and supports automated AI-driven public health responses. Future research should integrate real-time surveillance data and transmission dynamic models to further improve early warning precision.

**Keywords:** Seasonal influenza, Machine Learning, Early warning

## 1. Introduction

Seasonal influenza poses a severe threat to human health all over the world which continues to occur every year, with an estimated 1 billion people worldwide infected each year and 290,000-650,000 fatalities due to influenza-related respiratory diseases (1, 2).

The commencement of each influenza season is not fixed every year, as it can be influenced by viral survival, host immunity (3), population contact patterns (4), or viral interference caused by other respiratory pathogens (5). Therefore, it is crucial to timely signal any abnormal increase in the prevalence of influenza. Early detection and warning can reduce the harm to public health, providing a basis for the formulation of preventive measures and the preparation of medical resources. Previous studies have predominantly focused on predicting the trends of influenza outbreaks (6, 7). However, issuing timely warning signals holds greater practical significance for public health departments.

Various scenarios may trigger warning signals for influenza outbreaks, including surpassing historical baselines (8), excessively rapid increases in the prevalence trend, the discovery of new pathogens, higher disease severity, and inadequate medical resources (9). This study concentrates on the warning scenario where influenza enters an epidemic phase, characterized by an upward trend in the epidemic curve. In this scenario, the number of infected individuals increases, leading to a higher demand for medical resources. As the prevalence trend rises, the burden of the disease continuously intensifies, necessitating the implementation of specific personal protective or preventive measures to suppress the peak of the outbreak and slow down its spread. Issuing a warning at this stage does not imply the occurrence of a pandemic or a deviation from the usual trend but emphasizes the importance of health departments and society being attentive to the influenza epidemic.

Traditional early warning models primarily rely on binary classification methods (0/1 warnings), where a warning is issued only when a predefined threshold is exceeded. While these methods have been widely used in public health surveillance, they suffer from several critical limitations [10, 11]. First, binary warnings lack granularity, providing only a rigid "yes" or "no" decision rather than a nuanced risk assessment. This often leads to false alarms or missed detections, reducing the reliability of the system. Second, fixed-threshold models, such as the 40th percentile warning system, fail to distinguish between increasing and decreasing epidemic trends, sometimes issuing alerts even when cases are in decline. Lastly, these models do not allow for tiered public health responses, making it difficult for decision-makers to prioritize interventions based on varying levels of outbreak risk [12]. Addressing these limitations requires a more flexible and adaptive approach, where warning signals are expressed as continuous probability values rather than binary classifications [11, 12]. This study introduces a probability-based early warning model that enables tiered response strategies and AI-driven automation, offering a significant improvement over traditional methods.

The rise of machine learning and artificial intelligence methodologies provides a thrilling prospect for transforming influenza early warning systems. Traditional warning models, such as the MEM model (13), rely on historical data to predict the

current influenza prevalence. Machine learning methods, on the other hand, offer the potential to forecast warnings in advance. This shift towards advanced computational techniques is not just about improving the accuracy of predictions, it's about revolutionizing our ability to anticipate and respond to public health threats.

Machine learning models like the Self-Excitation Attention Residual Network (SEAR) have demonstrated the ability to dynamically adapt to evolving disease trends and surveillance system changes, enhancing the effectiveness of early warning systems [14]. These models can process complex, heterogeneous data from multiple sources, including influenza-like illness (ILI) cases, virological surveillance, climate and demography data, and even search engine queries. By doing so, they can capture intricate patterns and relationships that traditional models might miss, leading to more sensitive and accurate predictions.

Moreover, machine learning approaches have shown to outperform traditional models in terms of sensitivity, specificity, and timeliness. For instance, studies have found that machine learning models can issue warnings before the onset of influenza outbreaks with higher accuracy compared to threshold-based methods [15]. This is crucial for enabling public health officials to take preemptive measures, such as vaccination campaigns or public advisories, to mitigate the impact of the disease.

The potential of machine learning is further highlighted by the ability to integrate various data sources and apply complex nonlinear models like LSTM or other neural networks to the dataset, which can explore better prediction accuracy. These models can learn from genetic sequences and associated metadata to predict antigenic properties of circulating influenza viruses, providing insights into influenza evolution and antigenic drift.

The integration of machine learning and artificial intelligence into influenza early warning systems represents a significant leap forward in our capacity to predict and prepare for influenza outbreaks. These advanced methodologies offer a more proactive, data-driven approach to public health surveillance, potentially saving lives and reducing the societal and economic burden of influenza.

The primary objective of this study was to devise an innovative approach for early warning systems targeting influenza-like cases. A supervised deep learning model, Dense ResNet, was specifically developed for this purpose. The model's training involved fitting the ILI multiply Positive label, enabling the early detection and warning of changes in the activity level of influenza-like cases. This departure from conventional methodologies underscores the transformative potential of machine learning, particularly in providing advanced capabilities for timely and proactive warnings in the context of influenza outbreaks.

## 2. Method

### 2.1 Data sources

Northern China, Southern China were selected as the study target. We derived proxy measures for community influenza virus activity, referred to ILI+ rates, by multiplying the influenza-like illness (ILI) percentage (ILI%) with the proportions of influenza-positive specimens (13) from the Chinese National Influenza Surveillance Network for the two regions. A total of 569 weekly reports were included in this

study. The standardized ILI+ for each region was presented (**Figure 1**).

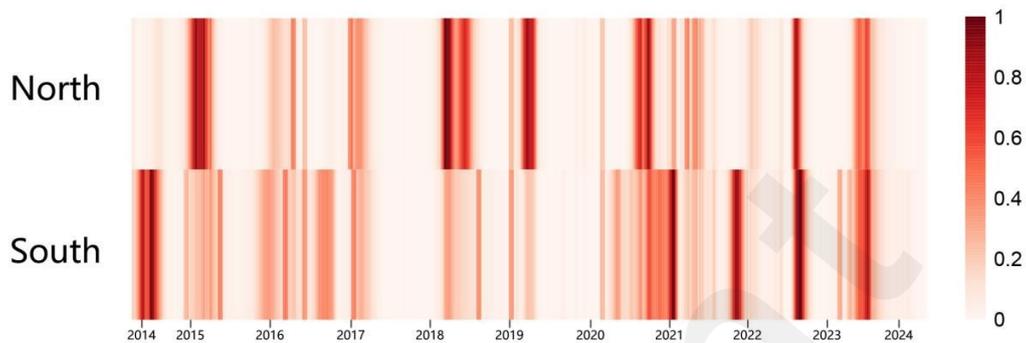
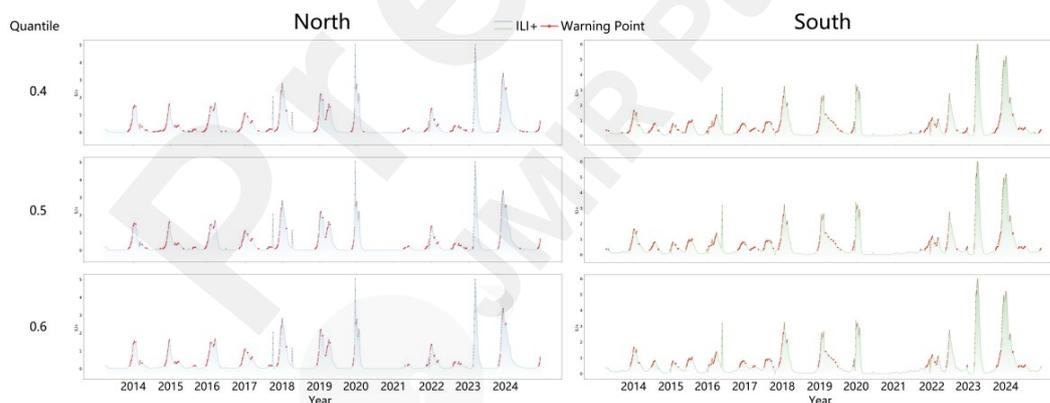


Figure 1. The ILI+ of Northern China and Southern China

## 2.2 Warning Definition and Labeling

Throughout the study, the warning scenario was defined as an ascending trend during the influenza season. Following the conventional definition of the influenza season in previous studies (17), we designated a warning when ILI+ exceeds the 40th percentile and demonstrates an upward trend. The labeled outcomes are illustrated in **Figure 2**.

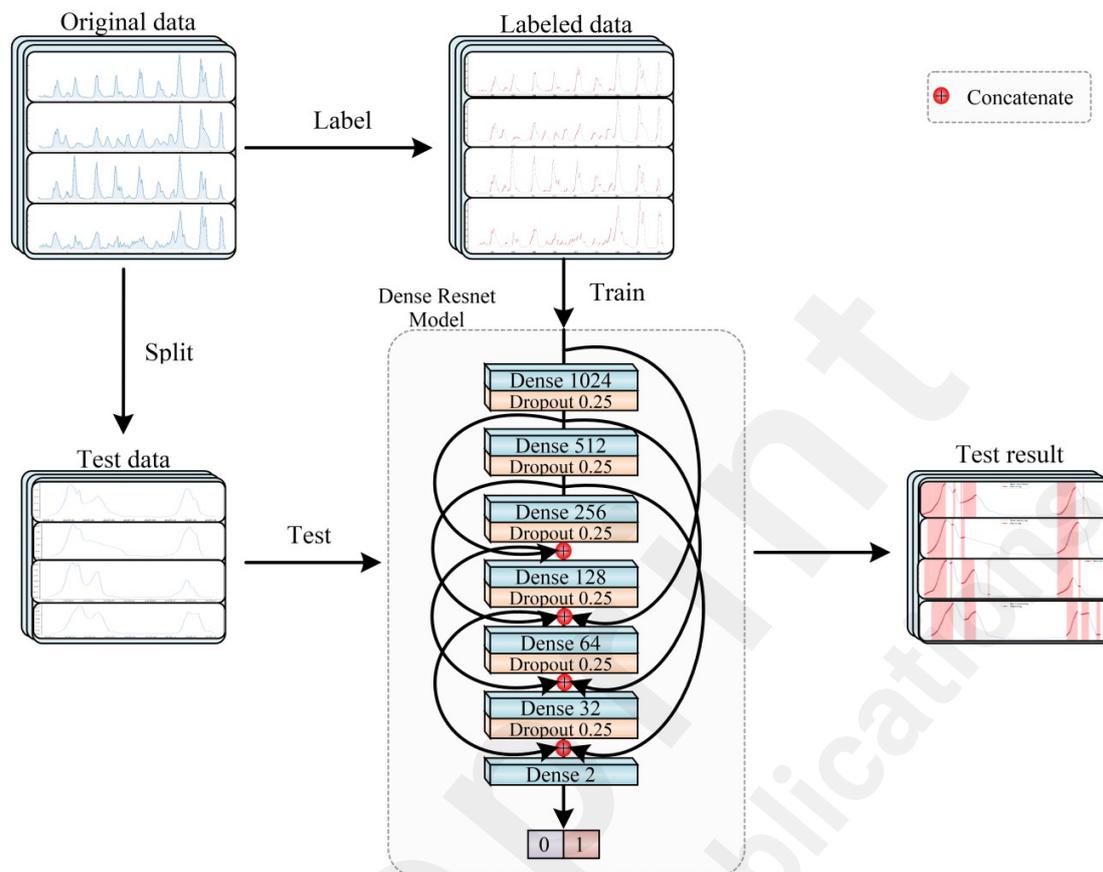
**Figure 2. ILI+ warning diagram for Northern and**



## **Southern China under different quantile values.**

### 2.3 Model construction

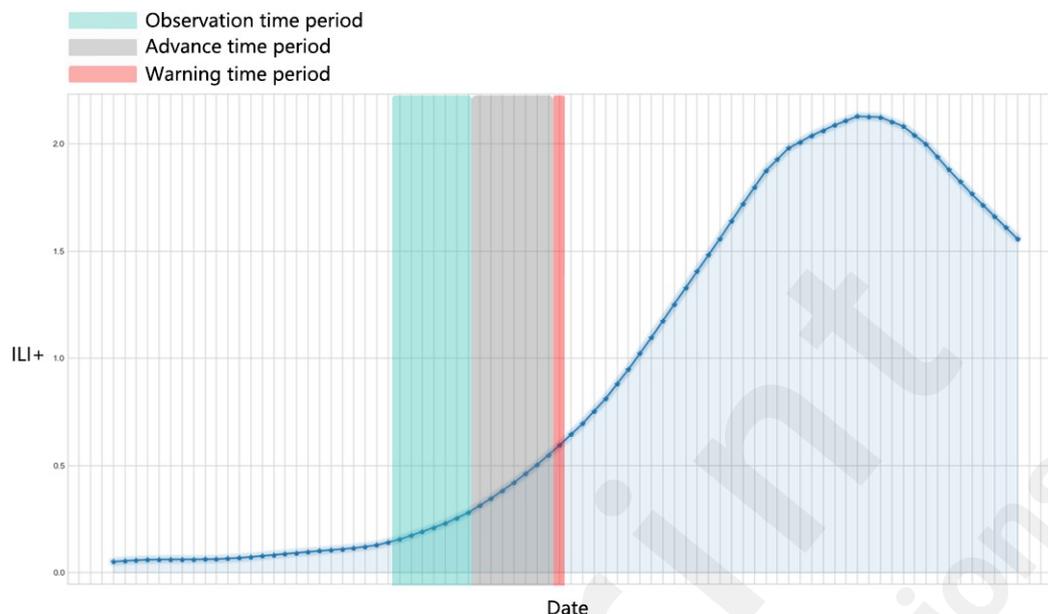
The objective of this study was to construct and train a model that could identify the intrinsic features of the input data and establish a mapping from ILI+ features to warning categories to accurately classify ILI+ for warnings. To achieve this, a deep learning model called Dense ResNet was innovatively proposed, as depicted in **Figure 3**.



**Figure 3. Structural diagram of Dense ResNet model**

In this study, to enhance the early warning function accuracy, Dense layers and the residual connection method were used to design the Dense ResNet model, as proven by experiments that the model could map ILI+ features to the early warning category effectively. The Dense ResNet model constructed in this study was composed of 7 Dense layers, 6 Dropout layers, and 7 residual connections. From top to bottom, Dense layers of 1,024, 512, 256, 128, 64, 32 units were employed to extract data features. Additionally, 6 layers with a dropout rate of 0.25 were applied after each Dense layer to prevent overfitting. Furthermore, 7 residual connections were added primarily to retain global information effectively while preventing gradient vanishing. The Sigmoid activation function was employed in the last fully connected layer to predict the probability of warning classes 0 and 1. The predicted result was a probability value between 0 and 1 for each warning category.

### 2.3 Dataset construction



**Figure 4. Schematic diagram of data construction**

To facilitate early warning, this study innovates on the construction of training data set. Figure 2 illustrates the dataset refactoring process graphically. We divided the data into three sections, namely observation time period and advance time period and warning time point. We combine the features of observation time period and the labels of warning time point to form a reconstructed dataset, and train this dataset to achieve early warning. For example, if you need to judge whether November 26, 2023 is a warning, you only need to observe the characteristic values from November 12, 2023 to November 18, 2023.

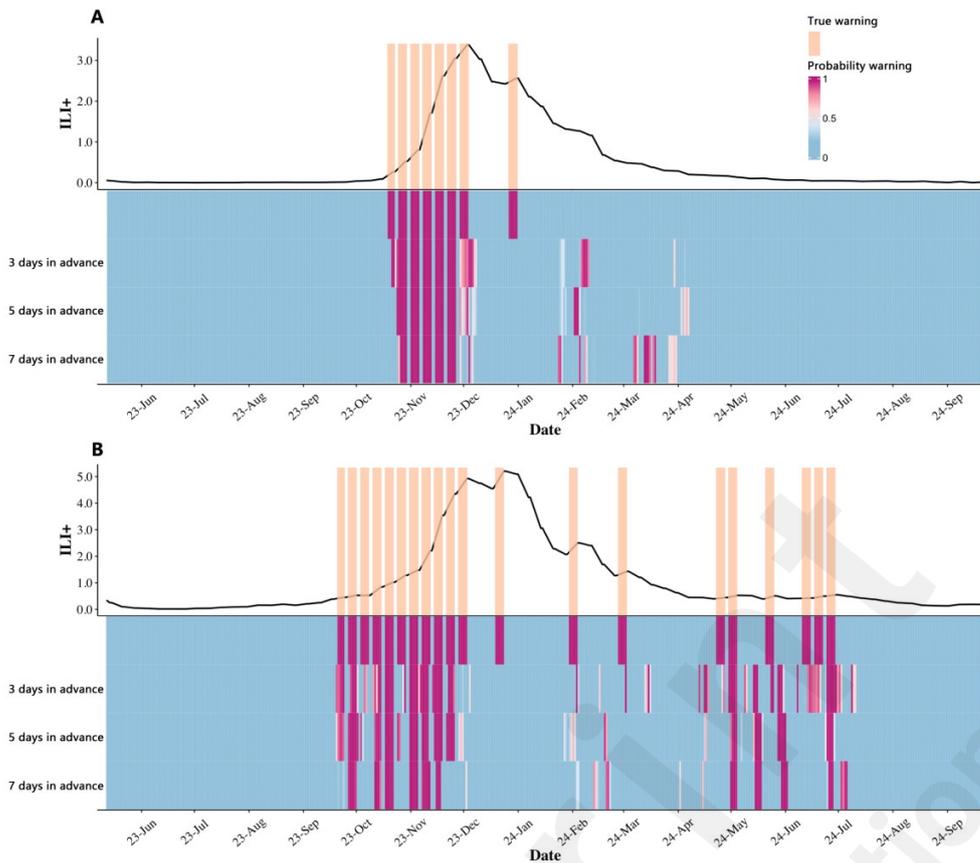
In this study, for the choice of advance time period, we adopted advance time periods of 3, 5, and 7 days, respectively, and obtained the results.

#### 2.4 Dataset segmentation

In this section, 3969 days spanning from December 30, 2013, to November 11, 2024, were selected in Southern and Northern China, to construct training. Following the conventional division methods of deep learning, including training set, validation set, and test set, and considering the practical significance of ILI% prediction, the period from December 30, 2013, to October 31, 2022, was designated as the training set, while the period from November 1, 2022, to November 11, 2024, was assigned as the test set. The training set encompasses 9 ILI% activity peaks, while the test set contains 2 ILI% peaks.

### 3. Results

Following the training and validation procedures employing the Dense ResNet model, early warning signals were issued at intervals of 3, 5, and 7 days in advance. The results indicate that, for both Northern and Southern China, the most effective lead time is 3 days, 50 quantile, allowing for the detection of warning signals 2 and 5



days earlier (**Figure 5**).

**Figure 5. Comparison of current/3,5,7 days in advance predicted warning singals in Northern and Southern China. A: Northern China; B: Southern China. Black line refers to ILI+ (%).**

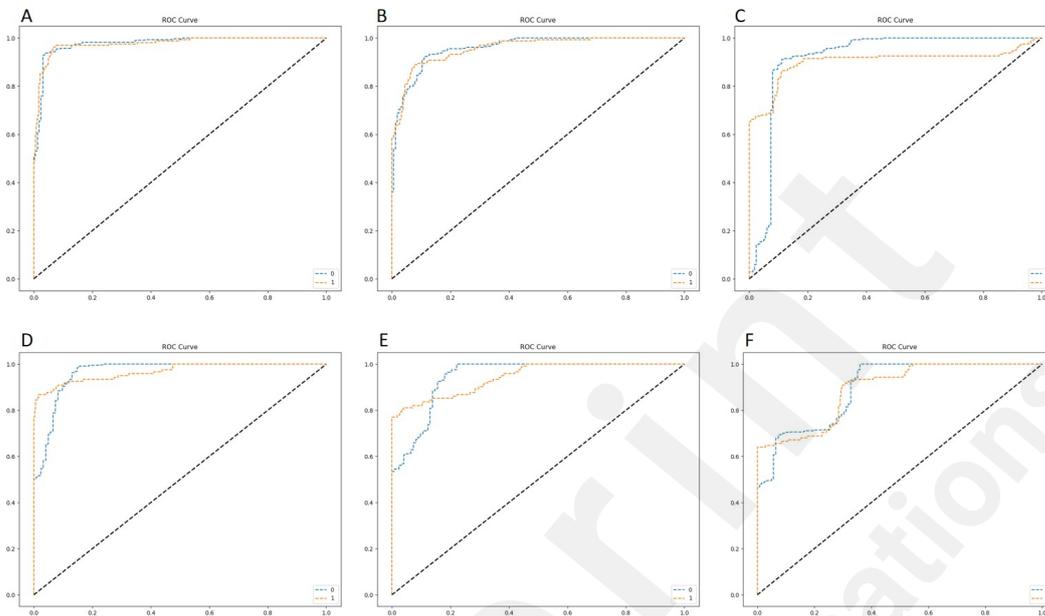
To validate the model, this study assessed warning performance in Northern and Southern China, utilizing metrics such as accuracy, recall, F1 score, and AUC. The outcomes highlight the highest AUC score for issuing warning signals 3 days in advance (Northern China: 0.94, Southern China: 0.95) (**Table 1**). The ROC curves are depicted in **Figure 6**.

In comparison with other commonly used models (support vector machine, random forest, xgboost, and LSTM), the Dense ResNet model exhibits superior warning performance in Southern and Northern China, as well as in Beijing and Yunnan province (**Table 2**).

**Table 1. Evaluation of early warning in Northern China, Southern China**

Days in advance	Warning Category	Precision	Recall	F1 score	AUC score
Northern China					
3	0	0.96	0.95	0.96	0.94
	1	0.89	0.93	0.91	
5	0	0.91	0.95	0.93	0.90
	1	0.88	0.81	0.84	
7	0	0.84	0.99	0.91	0.87
	1	0.99	0.60	0.75	
Southern China					
3	0	0.94	0.99	0.97	0.95
	1	0.98	0.81	0.89	
5	0	0.91	0.99	0.95	0.92

		1	0.99	0.69	0.82	
7	0	0.89	0.99	0.94	0.91	
	1	0.99	0.61	0.76		



**Figure 6. ROC curve of early warning effect in northern and southern China. A:** Northern China with 3 days in advance; **B:** Northern China with 5 days in advance; **C:** Northern China with 7 days in advance; **D:** Southern China with 3 days in advance; **E:** Southern China with 5 days in advance; **F:** Southern China with 7 days in advance.

**Table 2. Evaluation and comparison of 3 days early warning results of 5 models**

Model	Warning Category	Precision	Recall	F1 score	AUC score
<b>Northern China</b>					
SVM	0	0.86	0.95	0.90	0.81
	1	0.87	0.67	0.76	
XGboost	0	0.89	0.93	0.91	0.85
	1	0.85	0.76	0.80	
RF	0	0.88	0.90	0.89	0.85
	1	0.79	0.75	0.77	
LSTM	0	0.88	0.84	0.86	0.81
	1	0.69	0.75	0.72	
Dense Resnet	0	0.96	0.95	0.96	0.94
	1	0.89	0.93	0.91	
<b>Southern China</b>					
SVM	0	0.94	0.97	0.95	0.89
	1	0.89	0.80	0.84	
XGboost	0	0.93	0.87	0.90	0.84
	1	0.67	0.80	0.73	
RF	0	0.95	0.87	0.90	0.87
	1	0.68	0.85	0.76	
LSTM	0	0.89	0.93	0.91	0.87

	1	0.76	0.66	0.70	
Dense	0	0.94	0.99	0.97	
Resnet	1	0.98	0.81	0.89	0.95

SVM: Support Vector Machine

XGboost: Extreme Gradient Boosting

RF: Random Forest

LSTM: Long Short-Term Memory

## 4. Discussion

Seasonal influenza in China follows a predictable pattern, peaking during the colder months and, in some regions, during the summer. The ability to issue early and precise warnings is crucial for public health preparedness. This study introduces a continuous probability-based warning system (0-1 variable), moving beyond the traditional binary (0/1) warning models. This advancement significantly enhances the flexibility, accuracy, and applicability of early warning systems.

A probability-based risk assessment improves warning accuracy. Traditional early warning models rely on binary classifications—either issuing a warning or not—which often leads to false alarms or missed detections. These rigid threshold-based warnings do not account for the varying degrees of epidemic severity. In contrast, our continuous probability warning approach provides a more nuanced risk assessment, offering granular insights into the likelihood of an outbreak. This enables decision-makers to adjust intervention strategies based on the level of risk rather than adhering to a fixed binary outcome. As a result, the reliability and adaptability of early warning systems are significantly improved.

Moreover, a probability-based warning system supports tiered response strategies, which can be dynamically adjusted according to the predicted outbreak severity. For example, when the warning probability is moderate (0.4-0.6), internal monitoring may be prioritized, whereas higher probabilities ( $\geq 0.7$ ) would prompt targeted interventions such as increasing healthcare preparedness or reinforcing public health advisories. When the probability reaches 0.9 or above, immediate public health measures can be activated. This graduated approach prevents unnecessary disruptions in low-risk situations while ensuring timely action in high-risk scenarios.

A flexible warning system supports precise public health interventions. Another key advantage of probability-based warnings is their role in scientific policy-making and public communication. Traditional threshold-based methods may trigger unnecessary alarm in low-risk years, leading to economic losses and public fatigue toward warnings. Conversely, these methods may fail to escalate warnings quickly enough in high-risk years, delaying response measures. By shifting to a risk-probability framework, this study enables a more rational and transparent decision-making process, allowing authorities to communicate risk levels clearly and avoid overreactions.

Additionally, when the probability of an outbreak is low (e.g., 0.3-0.5), preventive measures can be targeted at high-risk groups, such as the elderly and individuals with pre-existing health conditions, without imposing large-scale interventions. This precision prevention strategy minimizes unnecessary economic

losses and social disruption while ensuring that those at the greatest risk receive appropriate protection.

AI-driven automation enhances the efficiency of warning responses. Beyond improving warning accuracy, continuous probability warnings pave the way for AI-driven automation in public health management. In future applications, AI-assisted decision-making systems can integrate probability forecasts into automated response frameworks as an example:

When the probability is  $\geq 0.7$ , automated alerts can be dispatched to healthcare facilities.

At  $\geq 0.9$  probability, public health advisories can be issued automatically to initiate broader containment measures.

For probabilities between 0.4 and 0.6, real-time monitoring can continue internally without triggering unnecessary public concerns.

By integrating machine learning and automation, this system improves response efficiency, reducing delays in intervention while optimizing the allocation of healthcare resources.

Early warning allows for timely interventions and improved outbreak control. The effectiveness of the proposed model was demonstrated through empirical validation. Our study shows that the machine learning model can issue warnings 5 days earlier than traditional methods, significantly improving public health preparedness. Given that preventive measures taken 5 days in advance can impact the infection chain by 2-4 generations, this early warning capability is crucial for controlling the spread of influenza. Similar findings have been observed in real-time surveillance systems for other infectious diseases, such as dengue, where early detection has been shown to reduce outbreak severity and enhance control efficiency.

Moreover, our results indicate that when predicting outbreaks 3 days in advance, the model achieves the highest sensitivity, specificity, and timeliness. This suggests that while earlier warnings provide more time for response, optimizing the trade-off between lead time and accuracy is essential. Future research should explore hybrid models that integrate transmission dynamics modeling and data-driven time-series forecasting to further improve warning precision.

Machine learning outperforms traditional threshold-based methods. Compared to the conventional fixed percentile warning system (e.g., 40th percentile thresholds), our probability-based approach provides three key advantages:

Earlier detection (5 days in advance): This allows for a crucial preparatory window for vaccine distribution, hospital readiness, and community interventions.

Distinguishing between rising and declining trends: Unlike fixed-threshold methods, which often issue warnings even after the peak, probability-based warnings dynamically adjust based on real-time trends.

Better adaptability to different epidemic conditions: The machine learning model learns from historical patterns, providing a flexible response mechanism even during atypical flu seasons.

A probability-based warning system enhances influenza preparedness. This study highlights the superior performance of probability-based machine learning models in seasonal influenza early warning.

## 5. Conclusion

This study presents a novel probability-based machine learning model for influenza early warning, demonstrating superior accuracy, flexibility, and practical applicability. By replacing binary warnings with probability-driven risk assessments, this approach enhances influenza preparedness and supports automated AI-driven public health responses. Future research should integrate real-time surveillance data and transmission dynamic models to further improve early warning precision. By issuing alerts 5 days earlier and providing a flexible, risk-adaptive approach, this system enhances public health preparedness and reduces unnecessary disruptions. Compared to traditional methods, machine learning accurately distinguishes epidemic trends, minimizes false alarms, and enables automated AI-driven public health responses. Future research should further refine warning frameworks by integrating transmission dynamics models and real-time surveillance data, ensuring even greater timeliness and effectiveness in influenza prevention and control.

## Declaration of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

1. This research was funded by: The Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences under Grants [2021-I2M-1-044, 2023-I2M-3-011]. National Key Research and Development Program of China [2023YFC2308701].

2. This research was supported by Biomedical High Performance Computing Platform, Chinese Academy of Medical Sciences.

## References

- [1] Iuliano AD, Roguski KM, Chang HH, et al. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study[J]. *Lancet*, 2018, 391(10127): 1285-1300. DOI: 10.1016/s0140-6736(17)33293-2.
- [2] Geneva: World Health Organization[EB/OL]. [2023-03-12]. [https://www.who.int/fr/news-room/fact-sheets/detail/influenza-\(seasonal\)](https://www.who.int/fr/news-room/fact-sheets/detail/influenza-(seasonal)).
- [3] Leung NHL. Transmissibility and transmission of respiratory viruses[J]. *Nat Rev Microbiol*. 2021;19(8):528-545. DOI:10.1038/s41579-021-00535-6
- [4] Feng, L., Zhang, T., Wang, Q. et al. Impact of COVID-19 outbreaks and interventions on influenza in China and the United States. *Nat Commun* 12, 3249 (2021). <https://doi.org/10.1038/s41467-021-23440-1>
- [5] Han S, Zhang T, Lyu Y, et al. The Incoming Influenza Season - China, the United Kingdom, and the United States, 2021-2022. *China CDC Wkly*. 2021;3(49):1039-1045. doi:10.46234/ccdcw2021.253
- [6] Ali ST, Cowling BJ. Influenza Virus: Tracking, Predicting, and Forecasting[J]. *Annu Rev Public Health*. 2021;42:43-57. doi:10.1146/annurev-publhealth-010720-021049
- [7] Venkatramanan S, Sadilek A, Fadikar A, et al. Forecasting influenza activity using machine-learned mobility map. *Nat Commun*. 2021;12(1):726. Published 2021 Feb 9. doi:10.1038/s41467-021-21018-5
- [8] Wu F, Kelsey A. Early Detection of Influenza Activity Using Syndromic Surveillance in Missouri[J]. *Online J Public Health Inform*. 2013;5(1):e37. Published 2013 Apr 4.
- [9] Zhang T, Wang Q, Leng Z, et al. A Scenario-Based Evaluation of COVID-19-Related Essential Clinical Resource Demands in China[J]. *Engineering (Beijing)*. 2021;7(7):948-957. doi:10.1016/j.eng.2021.03.020
- [10] Patel A, Maruthananth K, Matharu N, Pinto AD, Hosseini B. Early Warning Systems for Acute Respiratory Infections: Scoping Review of Global Evidence. *JMIR Public Health Surveill*. 2024 Nov 7;10:e62641. doi: 10.2196/62641.
- [11] Hu WH, Sun HM, Wei YY, Hao YT. Global infectious disease early warning models: An updated review and lessons from the COVID-19 pandemic. *Infect Dis Model*. 2024 Dec 3;10(2):410-422. doi: 10.1016/j.idm.2024.12.001.
- [12] Haque S, Mengersen K, Barr I, Wang L, Yang W, Vardoulakis S, Bambrick H, Hu W. Towards development of functional climate-driven early warning systems for climate-sensitive infectious diseases: Statistical models and recommendations. *Environ Res*. 2024 May 15;249:118568. doi: 10.1016/j.envres.2024.118568.
- [13] Murray JLK, Marques DFP, Cameron RL, et al. Moving epidemic method (MEM) applied to virology data as a novel real time tool to predict peak in seasonal influenza healthcare utilisation[J]. *Euro Surveill*. 2018;23(11):18-00079. doi:10.2807/1560-7917.ES.2018.23.11.18-00079
- [14] Yang L, Yang J, He Y, Zhang M, Han X, Hu X, Li W, Zhang T, Yang W. Enhancing infectious diseases early warning: A deep learning approach for influenza surveillance in China. *Prev Med Rep*. 2024 May 15;43:102761. doi: 10.1016/j.pmedr.2024.102761.
- [15] Kang SK, Son WS, Kim BI. Application of the Time Derivative (TD) Method for Early Alert

of Influenza Epidemics. *J Korean Med Sci.* 2024 Jan 29;39(4):e40.

doi: 10.3346/jkms.2024.39.e40.

[16] Goldstein, E., Cobey, S., Takahashi, S., Miller, J.C., Lipsitch, M., 2011. Predicting the epidemic sizes of influenza A/H1N1, A/H3N2, and B: a statistical method. *PLoS Med.* 8(7), e1001051.

[17] Du Z, Shao Z, Zhang X, et al. Nowcasting and Forecasting Seasonal Influenza Epidemics - China, 2022-2023. *China CDC Wkly.* 2023;5(49):1100-1106. doi:10.46234/ccdcw2023.206

[18] Alexander, R., & Alexander, M. (2014). An ICT-Based Real-Time Surveillance System for Controlling Dengue in Sri Lanka. *ArXiv*, abs/1405.4092.

[19] Lai, S., Ruktanonchai, N.W., Zhou, L. et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* 585, 410–413 (2020). <https://doi.org/10.1038/s41586-020-2293-x>

[20] Yang L, Zhang T, Han X, et al. Influenza Epidemic Trend Surveillance and Prediction Based on Search Engine Data: Deep Learning Model Study. *J Med Internet Res.* 2023;25:e45085.

Published 2023 Oct 17. doi:10.2196/45085