scientific reports

Check for updates

OPEN

Spatio-temporal epidemic forecasting using mobility data with LSTM networks and attention mechanism

Shihu Jiao[1], Yu Wang[1], Xiucai Ye[1]✉, Larry Nagahara[2] & Tetsuya Sakurai[1]

The outbreak of infectious diseases can have profound impacts on socio-economic balances globally. Accurate short-term forecasting of infectious diseases is crucial for policymakers and healthcare systems. This study proposes a novel deep learning approach for short-term forecasting of infectious disease trends, using COVID-19 confirmed cases and hospitalizations in Japan as a case study. This method provides weekly updates and forecasts outcomes over 1–4 weeks. The proposed model combines long short-term memory (LSTM) networks and multi-head attention mechanism strengths and is trained on public data sourced from open-access platforms. We conduct a comprehensive and rigorous evaluation of the performance of our model. We assess its weekly predictive capabilities over a long period of time by employing multiple error metrics. Furthermore, we carefully explore how the performance of the model varies over time and across geographical locations. The results demonstrate that the proposed model outperforms baseline approaches, particularly in short-term forecasts, achieving lower error rates across multiple metrics. Additionally, the inclusion of mobility data improves the predictive accuracy of the model, especially for longer-term forecasts, by capturing spatio-temporal dynamics more effectively. The proposed model has the potential to assist in decision-making processes, help develop strategies for controlling the spread of infectious diseases, and mitigate the pandemic's impact.

Keywords Infectious disease, Mobility, LSTM, Attention, Forecasting

Since the early 21st century, the world has witnessed several outbreaks of infectious diseases, such as severe acute respiratory syndrome (SARS), Middle East respiratory syndrome coronavirus (MERS-CoV), and most recently coronavirus disease 2019 (COVID-19), all posing significant threats to human health and severely impacting global socio-economic structures[1–4]. These pandemics bring fear and uncertainty to the social sphere, leading to significant changes in human behavior and social dynamics. Measures like lockdowns, social distancing, and travel restrictions have reshaped how people interact, work, and engage in social activities[5]. Additionally, hospitals are facing overwhelming numbers of patients, shortages of medical supplies and exhausted medical staff. Meanwhile, prioritizing the management of these diseases leads to deprioritization or delay of noncommunicable disease healthcare services, such as routine checkups, elective surgeries, and disease prevention programs, potentially leading to long-term health consequences[6,7].

Given the profound impacts of infectious diseases, it becomes clear that developing predictive models to forecast the pandemic's progression is not just beneficial but essential[8,9]. By forecasting the spread of the virus, the number of cases, and hospitalizations, policymakers can effectively allocate resources, implement targeted interventions, and develop evidence-based policies[10]. Furthermore, the ability to forecast infectious disease trends aids in scenario planning and enhances preparedness for potential future outbreaks. Reliable forecasts facilitate timely decision-making, allowing for proactive measures to mitigate the socio-economic and public health consequences of the pandemic[11,12].

Researchers have applied a variety of methods to generate short-term infectious disease forecasts, ranging from traditional statistical methods to more advanced machine learning approaches. Traditional methods, such as auto-regressive integrated moving average (ARIMA)[13], have been widely used for capturing linear trends and seasonal patterns in time series data. However, these methods often face challenges when dealing with non-

[1]Department of Computer Science, University of Tsukuba, Tsukuba 3058577, Japan. [2]Department of Chemical and Biomolecular Engineering, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD 21218, USA. ✉email: yexiucai@cs.tsukuba.ac.jp

linear relationships and complex temporal dynamics. Alternative approaches include statistical models, such as decision trees[14] and deep learning frameworks[15]. These methods are flexible to incorporate various potential variables and are highly dependent on the quality and availability of input data. Many studies have also explored the use of ensemble models, which combine multiple methods to integrate predictions from various approaches[16]. However, these models often require substantial computational resources and can reduce the interpretability of individual model contributions. Among these techniques, long short-term memory (LSTM) networks stand out for their proficiency in processing time series data[17–19]. Nevertheless, faced with an unprecedented global challenge like COVID-19, these traditional epidemic prediction models have many limitations, especially in capturing and understanding the complex dynamics of how the epidemic changes over time in different regions[20–22]. Advanced techniques, such as Transformer architectures and attention mechanisms, could be further explored to develop more effective forecasting models. Furthermore, previous research has shown that human mobility is crucial in disease transmission dynamics, with mobility patterns serving both as an indicator of public activity and as a measure of the effectiveness of social distancing[23]. Despite its potential, there has been limited systematic research on integrating mobility data into these advanced deep learning-based forecasting frameworks. In this study, we introduce an advanced deep learning framework designed to more accurately capture the spatiotemporal variability of health outcomes during disease outbreaks, as shown in Fig. 1. Unlike traditional models, our approach integrates LSTM networks with a Transformer-based attention mechanism to better capture temporal dependencies in disease transmission. Additionally, we try to systematically assess the role of mobility data in enhancing forecasting accuracy of the proposed framework across different regions. Focusing on COVID-19 as a case study, our model offers forecasts for cases and hospitalizations across all prefectures in Japan. It integrates epidemiological data and mobility patterns to improve prediction accuracy, with forecasting windows ranging from 1 to 4 weeks. The results demonstrate that the integration of mobility as a variable for model training enhances predictive performance. We rigorously evaluated the model's performance using multiple error metrics, considering different periods and regions. This could assist policymakers in implementing timely actions to alleviate the pandemic's severity and more efficiently manage healthcare resources (such as medical personnel, ICU beds, and medical supplies) during crises[24].

Methods

Dataset description

Using daily time-series data from Japan, we focus on prefecture-level data modeling and analysis. We collected raw time series data across all 47 prefectures of Japan between December 6, 2020 and October 16, 2021 (Fig. S1 shows the map of Japan). We only include data up to the end of 2021 because the emergence and dominance of the Omicron variant significantly changed the dynamics of the pandemic. Additionally, many regions began to relax COVID-19 restrictions, and changes in reporting practices resulted in the discontinuation or inconsistency of high-quality key data for model training and evaluation. This study focuses on three variables: COVID-19 confirmed case data, hospitalization data, and Google mobility data. The case and hospitalization data were

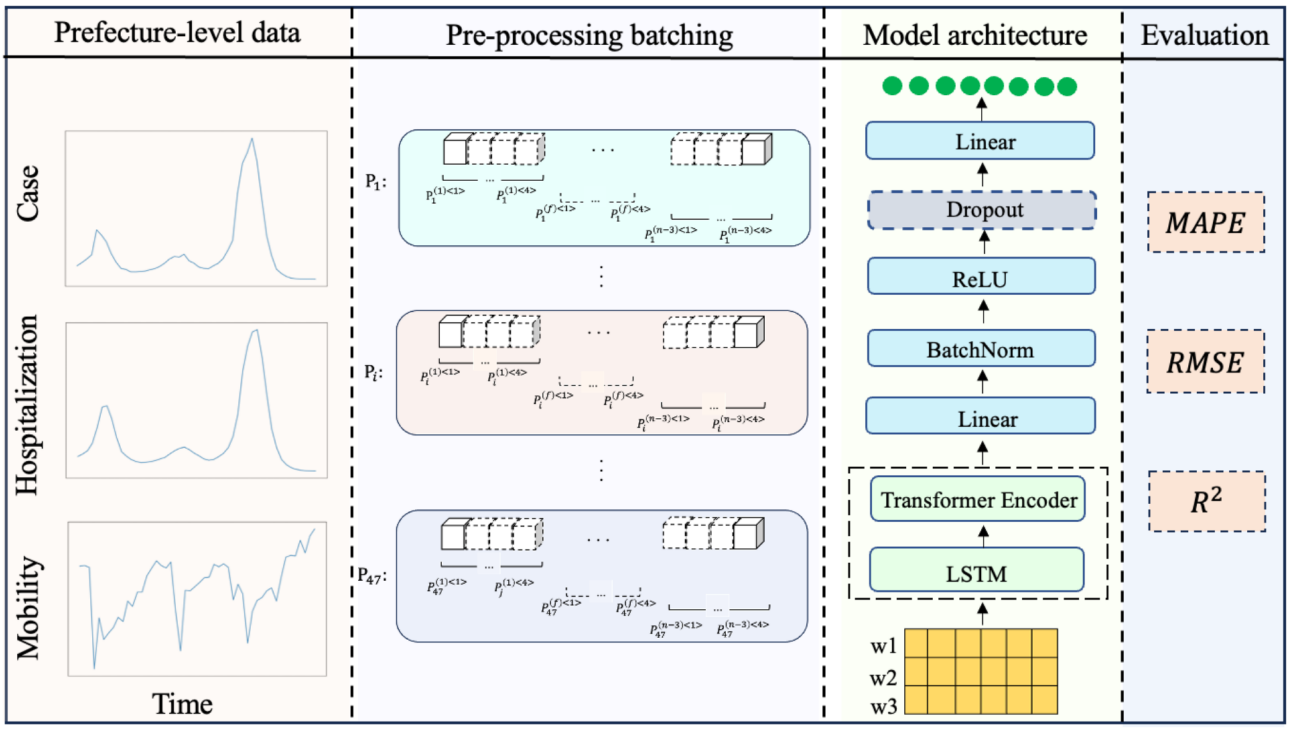

Fig. 1. Overall workflow of proposed deep learning based methods for estimating prefecture-level COVID-19 pandemic in Japan.

obtained by collecting the daily count of confirmed cases and daily count of COVID-19 patients admitted to hospitals in each prefecture from the Ministry of Health, Labour and Welfare (MHLW) (https://www.mhlw.go.jp/stf/covid-19/open-data.html). Additionally, the proposed model considers public mobility information, which offers insights into people's movement patterns in different locations. The Google mobility reports will be used to extract mobility statistics categorized by location and type of place, including "retail & recreation," "grocery & pharmacy," "parks," "transit stations," "workplaces," and "residential." Mobility data were obtained from Google mobility reports (https://www.google.com/covid19/mobility). These categories were treated as input features and integrated with the confirmed case and hospitalization data to capture the relationship between mobility patterns and COVID-19 dynamics.

Based on the epidemiological calendar, we computed the total number of confirmed cases and hospitalizations in all prefectures each week, starting on Sunday and ending on Saturday. We chose not to use daily data due to its higher noise level than weekly aggregated data. For each prefecture, we have 45 weeks of data. The final 12 weeks of the time series for all prefectures, covering the entire last pandemic wave, were reserved for testing and not used during training. Consequently, our model was trained using the initial 33 weeks of data. As in the previous study, we utilized a window size of four to create the feature vectors and labels for each time series[21].We then shuffled the feature vectors derived from the time series of all sub-populations, generating a comprehensive set of feature vectors. Next, we randomly allocated 70% of these vectors for training while reserving the remaining 30% for validation purposes.

Model overview

Previous research has shown that the transmission patterns of infectious diseases, such as COVID-19, are complex and dynamic over time, making accurate forecasting of disease dynamics a challenging task. To address this, developing a robust predictive modeling framework that considers the intricate interactions between various factors influencing transmission dynamics is crucial. Additionally, the framework must consider the uncertainty surrounding the timing and impact of these factors on observed transmission dynamics. We can enhance the accuracy and reliability of near-term disease forecasting by carefully selecting relevant input data streams and employing advanced modeling techniques. Given these challenges, we present a novel multi-stage deep learning framework. This framework leverages the initial first-stage prediction, based on the past three weeks of data, to forecast the subsequent week and iteratively applies this approach to predict further into the future. This paper focuses on 4-stage horizon forecasting (e.g., four-week ahead forecast). The framework is designed to predict three key variables: confirmed cases, hospitalizations, and mobility data.

As shown in Fig. 1, the model architecture begins with an LSTM layer, which has a layer with 64 neurons. The LSTM layer utilizes long short-term memory units to capture temporal dependencies in the input data. These units enable the model to learn patterns and trends in the time series data. Following the LSTM layer, we utilize a transformer encoder layer. The transformer encoder is responsible for attending to and aggregating relevant information from the input sequence. The transformer architecture is based on self-attention mechanisms, which allow the model to capture global dependencies in the data. This layer helps the model capture complex interactions between the factors affecting transmission dynamics. After the transformer encoder layer, there is a fully connected layer with 32 neurons. This layer serves as an intermediate step before the final prediction. It allows the model to process the encoded representations further and extract relevant features. We include a dropout layer after the fully connected layer to avoid overfitting. This layer helps prevent the model from relying too heavily on certain input features, promoting more robust generalization and preventing overfitting. The models were implemented using the PyTorch 2.2.0 package. The coding process took place in the Google Colaboratory environment, which provided GPU accelerators and a high-RAM runtime shape to enhance computational efficiency. A batch size of 512 was used to train the models, along with the Adam optimizer with a learning rate of 0.001. The models were trained for 300 epochs, utilizing the Huber loss function as the optimization criterion.

Model evaluation

In our comprehensive assessment of the model's performance, we employed a rigorous evaluation framework. Our evaluation spans a significant period of 12 weeks from July 25 2021 to October 16 2021, allowing us to capture a wide range of epidemiological weeks and assess the model's performance over different pandemic stages. The model generates predictions for a 4-week horizon at each evaluation step using an autoregressive approach, where previously predicted values are iteratively used as inputs for subsequent predictions. After each 4-week forecast, the input data is shifted forward by one week, and the process repeats. We then collected all the predictions for the first, second, third, and fourth weeks separately and compute the evaluation metrics accordingly. As a result, the 12-week evaluation period can generate 9 evaluation steps. For the assessment, we mainly used the mean absolute percentage error (MAPE), which is commonly used in related studies. However, MAPE has some limitations. When the true value is close to zero, MAPE will produce disproportionately large and unstable errors. In addition, MAPE is sensitive to extreme values and noise, which may distort the results, especially in datasets with significant outliers or small values. Therefore, we also employed two other commonly used metrics: root mean squared error (RMSE) and coefficient of determination (R^2). RMSE estimates the average squared difference between the predicted and actual values, measuring the consistency of model predictions with observations. On the other hand, R^2 evaluates how well our model fits the data, and the closer the value is to 1, the better the model performance. All of these metrics offer different perspectives on the model's performance and help us evaluate its accuracy, reliability, and overall fit effectiveness. The calculation formulas for these evaluation metrics are as follows:

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(A_i - P_i)^2} \tag{1}$$

$$\text{MAPE} = \frac{1}{T}\sum_{i=1}^{T}\left|\frac{A_i - P_i}{A_i}\right| \tag{2}$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{3}$$

where T is the sample size, A_i is the $i\,th$ actual value, and P_i is the $i\,th$ predicted value. Additionally, SS_{res} is the residual sum of squares and SS_{tot} is the total sum of squares.

Results

Comparison of model architectures

Our first goal was to assess the influence of incorporating mobility data on prediction performance. To address this question, we developed two models with distinct specifications. Model 1 utilized time series data for two variables: new confirmed cases and hospitalizations. In contrast, Model 2 incorporated additional features derived from mobility data. Figure 2 compares the box plots of MAPE and R^2 distributions for the two models over four forecast horizons (i.e., 1–4 weeks). A comparison of RMSE values is also provided in Supplementary Figure S2. Regarding MAPE, Model 1 and Model 2 exhibit similar trends in predicting both cases and hospitalizations. It is evident that both methods perform well on shorter time frames; however, as the prediction period extends, there is a noticeable increase in the MAPE values, indicating a decline in model performance. For the first week of forecasting, Model 1 performed slightly better than Model 2. For the following three weeks, it can be observed that Model 2, which incorporates mobility as a predictor, achieves a lower median and variance

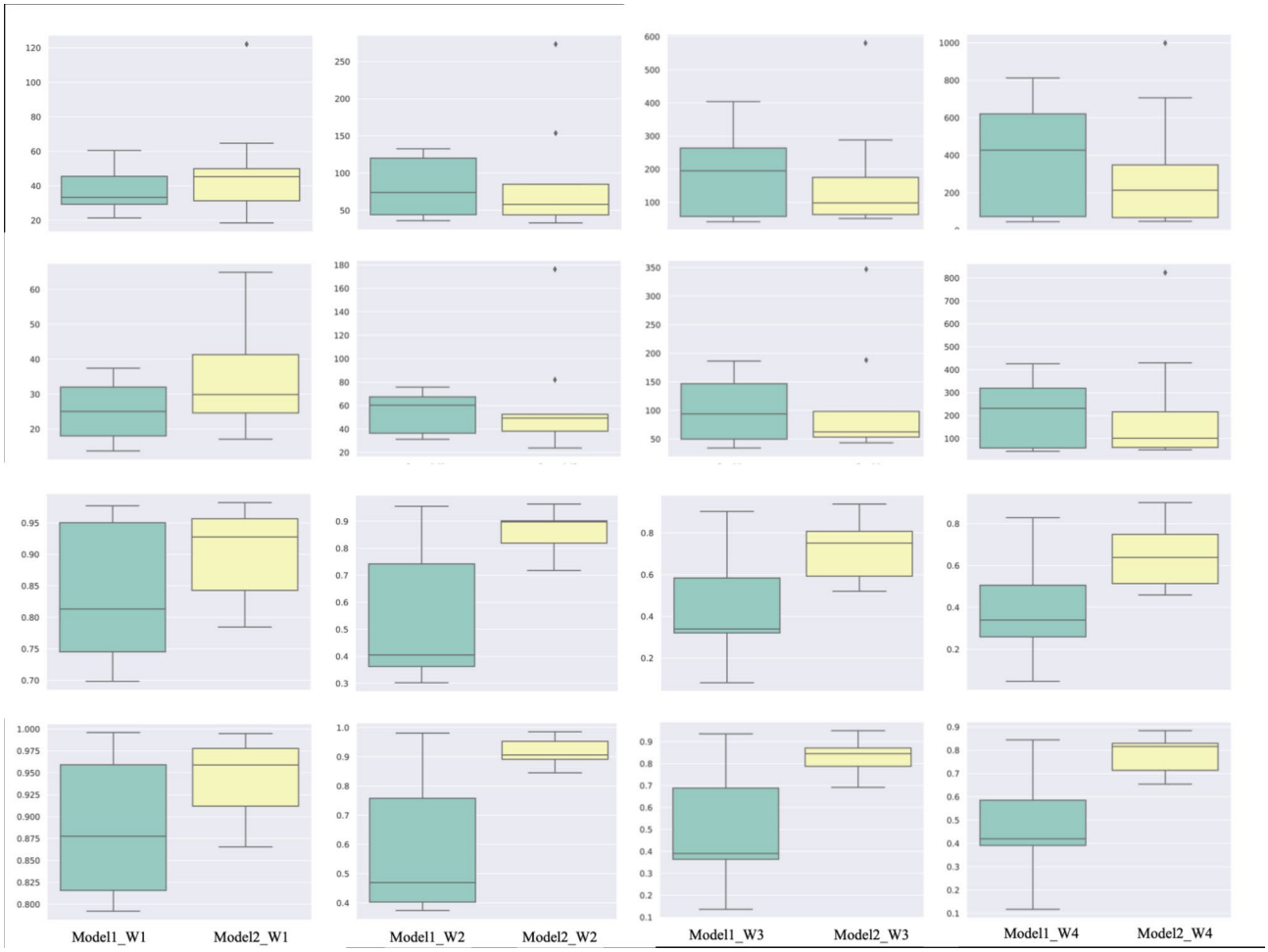

Fig. 2. **A** and **C** compare the performance of Model 1 and Model 2 in cases prediction using MAPE and R^2, respectively. **B** and **D** make a similar comparison for hospitalization prediction. The green bars are the error distributions for Model 1 (without mobility as a predictor), and the yellow bar are the error distributions for the Model 2 (with mobility as a predictor). The y-axis ranges differ across the plots to ensure the clarity and distinguishability of the data distribution.

in MAPE for predicting cases and hospitalizations compared to Model 1, which does not consider mobility. Including population mobility in the model enhances its performance, particularly when forecasting over a longer time horizon. This trend is even more pronounced for the box plot of R^2. Model 2 demonstrates more stable performance over time, demonstrating its robustness to variations across different periods. Model 2 also consistently achieves lower RMSE scores across all weeks (Supplementary Figure S2), suggesting that its predictions are more accurate and that mobility data contributes positively to its performance. These observed improvements in prediction performance can be attributed to the additional insights from mobility data. To further validate the impact of mobility data on model performance, we conducted an additional experiment by systematically removing individual mobility features (e.g., 'retail and recreation', 'grocery and pharmacy', 'parks', 'transit stations', 'workplaces', and 'residential'). As shown in Figure S3, removing any feature generally led to an increase in MAPE, especially in longer-term forecasts. Therefore, our final model includes all mobility features.

To further evaluate the effectiveness of introducing the Transformer architecture into our model for pandemic prediction, we developed an additional model (Model 3) for comparison. Model 3 utilized two LSTM layers while keeping other configurations the same as Model 2. Table S1 presents the mean MAPE values for Models 2 and 3 over four forecast horizons. For case predictions, Model 2 achieved lower MAPE values, especially at longer forecast horizons. Similarly, for hospitalization predictions, Model 2 outperformed Model 3, showing lower error rates across all weeks. This indicates that the Transformer-based model provides more accurate predictions than the LSTM-based model, demonstrating its robustness in different forecasting scenarios. Although both models benefit from the inclusion of mobility data, the superior performance of Model 2 can be attributed to the capabilities of the Transformer architecture. Transformers are particularly adept at capturing long-term dependencies and complex interactions within the data due to their self-attention mechanism, which allows the model to focus on different parts of the data simultaneously. This makes Model 2 more effective in handling the temporal dynamics and intricate relationships that characterize pandemic spread.

Model performance across prefectures

The deep-learning-based forecasting methodology utilized in our study is constructed upon prefecture-level COVID-19 data, facilitating an investigation into spatial variations in predictive performance. Figure 3 illustrates the geographical distribution of MAPE values across different prefectures, showing the disparities in forecast accuracy. The accuracies are computed over four forecast horizons. The color scales indicate the magnitude of the error. The deeper color corresponds to a more significant error.

The performance of the predictive models for both COVID-19 cases and hospitalizations demonstrates specific common patterns across all prefectures. One common observation is the increasing MAPE over forecast horizons. On the one hand, when predictions are used instead of actual observations, the recursive approach can cause the accumulation of prediction errors. This can lead to a decline in performance, especially as the forecast period becomes longer. On the other hand, various factors, such as changes in the spread of the virus, evolving

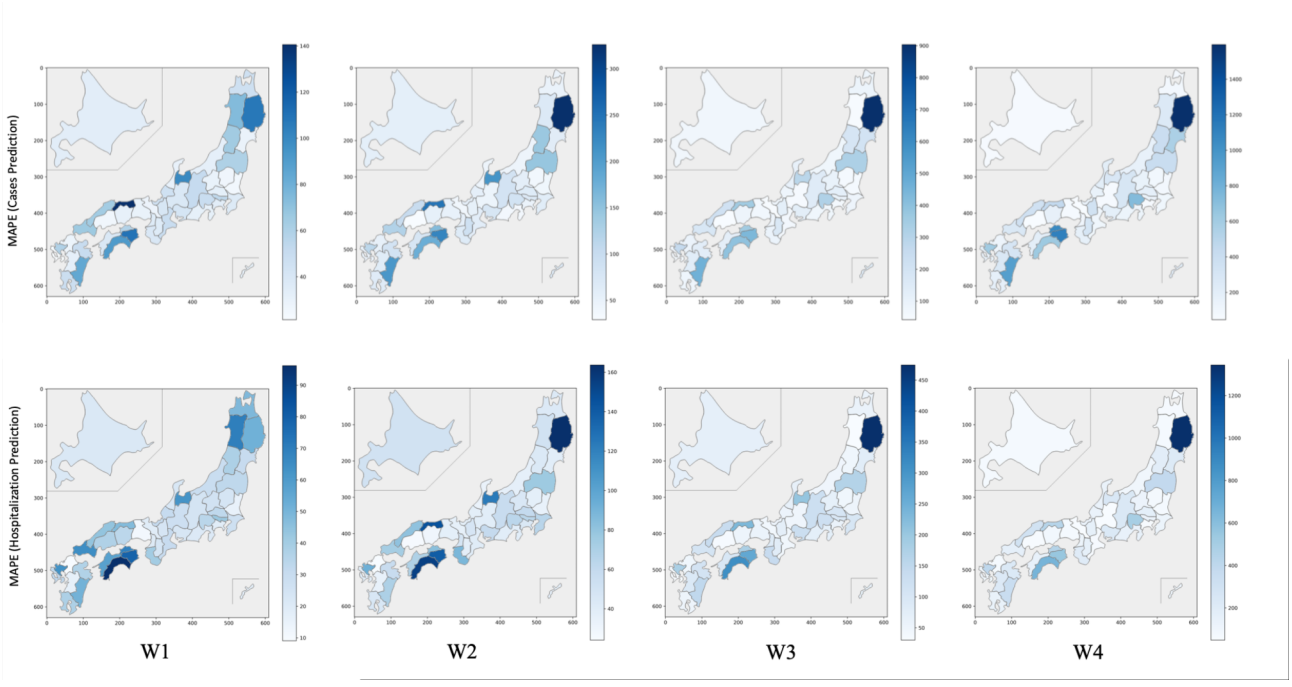

Fig. 3. **A** and **B** show the model performances for specific prefectures for one-to-four week ahead predictions of cases and hospitalizations, respectively, as measured by the MAPE. The color scales represent the magnitude of the error metric; the deeper color corresponds to larger error. The color scale ranges differ across the figures to ensure clarity and distinguishability of the error distribution. All maps were generated using the Japanmap package (version 0.0.24) [https://github.com/SaitoTsutomu/japanmap].

testing strategies, or unpredictable events, could also lead to a decrease in prediction accuracy over time. From a spatial perspective, the models' accuracy shows noticeable differences across prefectures. Some areas, namely Kanagawa, Osaka, Aichi, Saitama, and Chiba, consistently present lower MAPE and RMSE values, indicating more precise case and hospitalization predictions. On the contrary, other prefectures such as Miyagi, Iwate, Fukushima, Miyazaki, Tokushima, Yamanashi, Saga, Kagawa, and Kochi show higher error values, signifying lower prediction accuracy or increased variability. Iwate exhibits consistently high MAPE values across all weeks, indicating a significant discrepancy in both case and hospitalization predictions. Certain prefectures, including Miyagi, Fukushima, Miyazaki, and Kagawa, reveal a noticeable escalation in error values from one to four weeks ahead. This indicates a potential decline in the accuracy of case and hospitalization forecasts. Additionally, Yamaguchi, Nara, and Kagawa see a sharp rise in MAPE values between three-week ahead prediction and four-week ahead prediction, denoting a considerable drop in prediction accuracy. Interestingly, some prefectures, such as Toyama, exhibit a decrease in error values from a three-week ahead prediction to a four-week ahead prediction. This exciting trend deserves further investigation and analysis in future studies.

We also try to draw some conclusions about the model's performance based on population size. However, there are no clear spatial patterns and straightforward relationships between population size and MAPE variations (Figure S4). To further explore this, we divided the 47 prefectures into two groups using the 75th percentile of population size (population data from https://www.stat.go.jp/data/jinsui/2019np/index.html). The results, shown in Fig. 4, indicate that smaller population prefectures often have more outliers, especially in later stages (W3 and W4). This suggests that over time, smaller prefectures tend to show greater variability and higher prediction errors. In contrast, larger population prefectures exhibit more consistent model performance. A potential reason for the higher variability and prediction errors in smaller prefectures could be related to the noisier quality of mobility data in rural areas. Large metropolitan areas, which are well-monitored, often have more readily available and accurate foot traffic data, whereas rural areas may suffer from less comprehensive data coverage.

Visualizing the prediction results

To provide additional evidence of the forecasting capabilities of our models, we visualize the actual and predicted data for each prefecture in Supplementary Figures S3–S10. We segmented the data according to different forecast horizons, allowing for a more precise comparison between the predicted and actual trends. As an example, we display the predictions for Okayama Prefecture in Fig. 5. The actual numbers are in blue, while our deep learning model's forecasts are in orange. Figure 5A demonstrates the high accuracy of our one-week ahead case predictions, with minimal error observed between the predicted and actual case values. The overlapping of the predicted and actual values at some of the time points further reinforces the effectiveness of our proposed method. For the 2–4 week forecasts, the predicted line on the left of the peak deviates gradually from the actual value line. Moreover, the divergence between the predicted line and the actual values on the right side of the peak becomes significant in the fourth week. The prediction of hospitalization is similar but overall better than the case prediction (Fig. 5B), with the predicted values usually very close to the actual values. Our model's performance varies depending on the phase of the outbreak: it performs less accurately during periods of rapid increase, but its accuracy improves during periods of rapid decrease. This suggests that our model is more effective in capturing trends during the decreasing phases of the outbreak and needs enhanced capabilities during the increasing phases to better manage and respond to the crisis.

Discussion

The results of this study show that mobility data plays a crucial role in improving the predictive performance of epidemic forecasting models, especially for long-term predictions. This is likely because it captures movement patterns and population behavior that influence the spread and transmission of the virus, a finding that aligns with existing research. Our study introduces a novel deep learning framework that combines the strengths

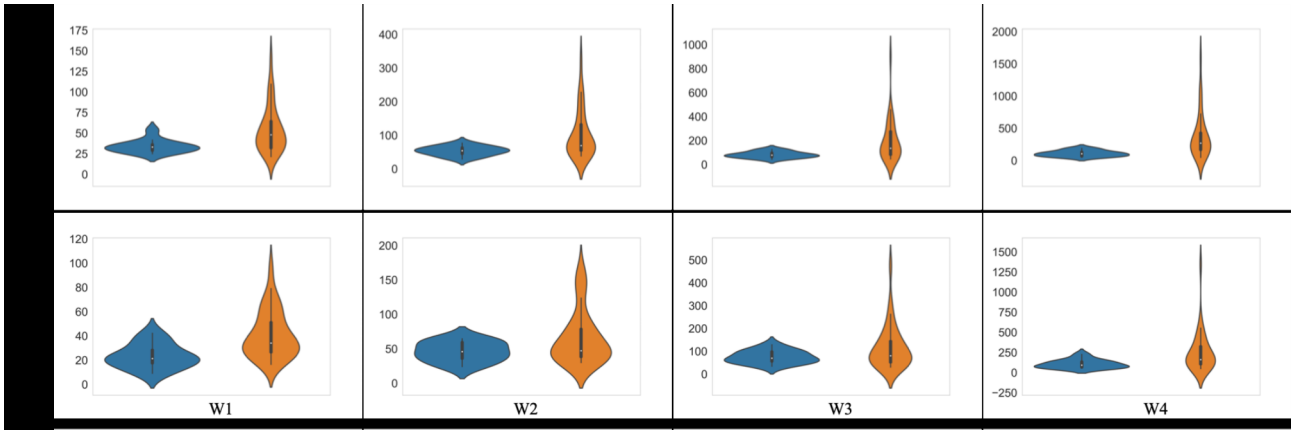

Fig. 4. Distribution of MAPE values for large (blue) and small (orange) population prefectures across four forecasting stages.

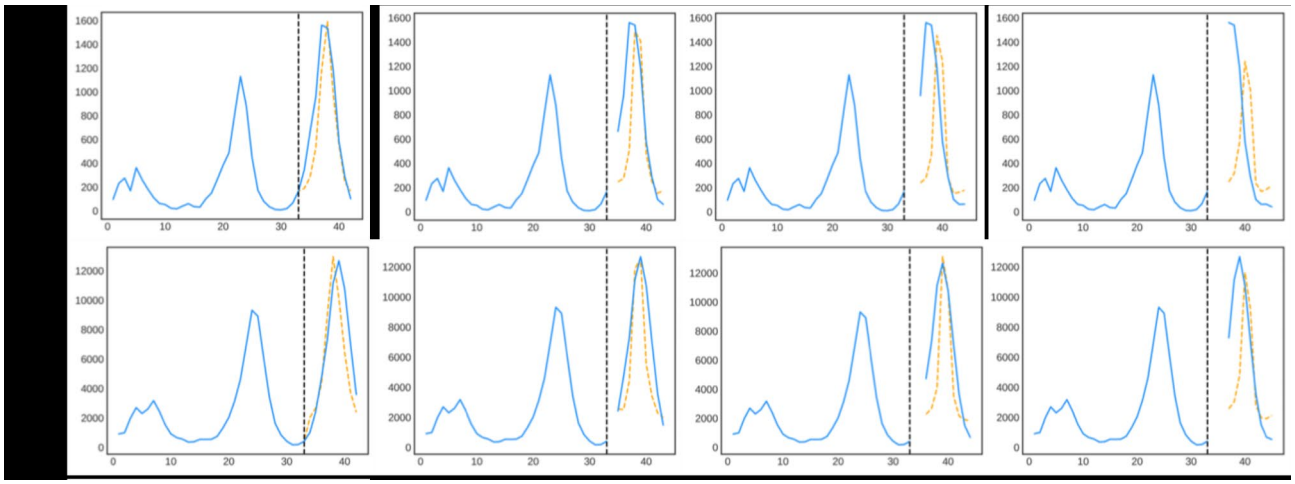

Fig. 5. Actual and forecasted number of cases (**A**) and hospitalizations (**B**) in Okayama for one-to-four week prediction windows.

of LSTM networks and Transformer architecture, enabling the model to capture complex dynamics patterns of infectious diseases with improved accuracy and robustness. Moreover, our model demonstrates potential for real-world applications, such as helping policymakers and health authorities make informed decisions, including the implementation of targeted interventions and preventive measures in areas with higher mobility and potential disease transmission.

Despite its promising applications, the model still has limitations, and its performance needs further improvement. First, errors can accumulate in multi-step forecasting, and the model relies heavily on historical data, which can be problematic during outbreak peaks or sudden declines. Second, many prefectures in Japan have relatively small populations, leading to low case and hospitalization numbers during certain periods. As a result, even small absolute errors can translate into disproportionately high relative errors. Additionally, these regions often have noisier and less comprehensive mobility data, show more variability and higher prediction errors compared to larger, well-monitored regions. Third, the inherent noise and inconsistencies in epidemiological and mobility data, particularly those influenced by government interventions, seasonality, and public behavior, further exacerbate prediction errors. Finally, potential underreporting is also an important issue, which can negatively impact model accuracy, particularly in small prefectures.

Based on these facts, several key areas can be further explored in the future. One key aspect is that, although our model demonstrates robust performance within the selected evaluation period, its stability over a longer timeframe still needs further evaluation. Another important direction is to improve the quality and completeness of data in underrepresented areas by developing better data collection methods and integrating other data sources, such as social media activity or traffic usage statistics. In addition, more factors should be considered, such as the local dynamics of the epidemic, population density, geographic location, and the effectiveness of containment measures implemented by regions. Analyzing these additional factors would provide a deeper understanding of disease spread and make the model more robust. Finally, it is important to study how virus transmission methods and evolutionary changes affect disease spread. Understanding how different transmission pathways and viral variants influence the epidemic can help the model adapt to new threats and changing situations. By addressing these areas, future research can build on the current study to develop more resilient and accurate epidemic forecasting models.

Conclusion

The emergence and spread of infectious diseases highlight the urgent need for innovative methods to predict their trajectories to maintain socioeconomic stability and public health. This study successfully demonstrates the application of novel deep learning models to short-term predictions of infectious disease trends, with a specific focus on COVID-19 cases and hospitalizations across Japan. By combining the power of LSTM and Transformer and leveraging mobility data, our model enhances forecasting performance for disease spread, providing weekly updates with a prediction window ranging from 1 to 4 weeks. Our comprehensive evaluation using multiple metrics highlights the model's robustness within the analyzed timeframe, as well as its reliability in capturing the temporal and spatial dynamics of infectious disease spread. The ability to offer detailed geographic forecasts can provide a nuanced understanding of disease spread in different regions, thereby facilitating targeted and effective interventions by policymakers and healthcare systems.

Data availability

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Received: 25 December 2024; Accepted: 11 March 2025

Published online: 20 March 2025

References

1. Hawryluck, L. et al. SARS control and psychological effects of quarantine, Toronto, Canada. *Emerg. Infect. Dis.* **10** (7), 1206 (2004).
2. Azhar, E. I. et al. The middle east respiratory syndrome (MERS). *Infect. Dis. Clin.* **33** (4), 891–905 (2019).
3. Martin, A. et al. Socio-economic impacts of COVID-19 on household consumption and poverty. *Econ. Disasters Clim. Change* **4** (3), 453–479 (2020).
4. Gagnon, J. E., Kamin, S. B. & Kearns, J. The impact of the COVID-19 pandemic on global GDP growth. *J. Jpn. Int. Econ.* **68**, 101258 (2023).
5. Bavel, J. J. V. et al. Using social and behavioural science to support COVID-19 pandemic response. *Nat. Hum. Behav.* **4** (5), 460–471 (2020).
6. Ranney, M. L., Griffeth, V. & Jha, A. K. Critical supply shortages — the need for ventilators and personal protective equipment during the Covid-19 pandemic. *N. Engl. J. Med.* **382** (18), e41 (2020).
7. Maringe, C. et al. The impact of the COVID-19 pandemic on cancer deaths due to delays in diagnosis in England, UK: a National, population-based, modelling study. *Lancet Oncol.* **21** (8), 1023–1034 (2020).
8. Viboud, C. & Vespignani, A. The future of influenza forecasts. *Proc. Natl. Acad. Sci.* **116** (8), 2802–2804. (2019).
9. Petropoulos, F. & Makridakis, S. Forecasting the novel coronavirus COVID-19. *PloS ONE* **15** (3), e0231236 (2020).
10. Holmdahl, I. & Buckee, C. Wrong but useful—what covid-19 epidemiologic models can and cannot tell us. *N. Engl. J. Med.* **383** (4), 303–305 (2020).
11. Huppert, A. & Katriel, G. Mathematical modelling and prediction in infectious disease epidemiology. *Clin. Microbiol. Infect.* **19** (11), 999–1005 (2013).
12. Johansson, M. A. et al. An open challenge to advance probabilistic forecasting for dengue epidemics. *Proc. Natl. Acad. Sci.* **116** (48), 24268–24274 (2019).
13. Benvenuto, D. et al. Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data Brief* **29**, 105340 (2020).
14. Watson, G. L. et al. Pandemic velocity: forecasting COVID-19 in the US with a machine learning & bayesian time series compartmental model. *PLoS Comput. Biol.* **17** (3), e1008837 (2021).
15. Ramchandani, A., Fan, C. & Mostafavi, A. Deepcovidnet: an interpretable deep learning model for predictive surveillance of covid-19 using heterogeneous features and their interactions. *Ieee Access* **8**, 159915–159930 (2020).
16. Cramer, E. Y. et al. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the US. *Medrxiv* (2021).
17. Chimmula, V. K. R. & Zhang, L. *Time Series Forecasting of COVID-19 Transmission in Canada Using LSTM Networks* 135 (Chaos Solitons & Fractals, 2020).
18. Zeroual, A. et al. *Deep Learning Methods for Forecasting COVID-19 time-Series Data: A Comparative Study* 140 (Chaos Solitons & Fractals, 2020).
19. Li, A. & Yadav, N. An adaptable LSTM network predicting COVID-19 occurrence using time series data. In *IEEE International Conference on Digital Health (ICDH) / IEEE World Congress on Services (SERVICES)*. (2021).
20. Rashed, E. A. & Hirata, A. One-Year lesson: machine learning prediction of COVID-19 positive cases with meteorological data and mobility estimate in Japan. *Int. J. Environ. Res. Public Health* **18** (11). (2021).
21. Rashed, E. A. & Hirata, A. Infectivity upsurge by COVID-19 viral variants in Japan: evidence from deep learning modeling. *Int. J. Environ. Res. Public Health* **18** (15). (2021).
22. Rashed, E. A., Kodera, S. & Hirata, A. COVID-19 forecasting using new viral variants and vaccination effectiveness models. *Comput. Biol. Med.* 149. (2022).
23. Nikparvar, B. et al. Spatio-temporal prediction of the COVID-19 pandemic in US counties: modeling with a deep LSTM neural network. *Sci. Rep.* **11** (1), 21715 (2021).
24. Shanbehzadeh, M. et al. Predictive modeling for COVID-19 readmission risk using machine learning algorithms. *BMC Med. Inf. Decis. Mak.* **22** (1), 139 (2022).

Acknowledgements

The authors thank Dr. Anton Dahbura for his helpful suggestions and comments, which have improved the quality of the article.

Author contributions

T.S. and L.N. conceived and designed this paper. T.S. and X.Y. offered advanced suggestions about the paper. S.J. and Y.W. analyzed the data. S.J., X.Y. and Y.W. wrote the draft of the paper. All authors reviewed the manuscript.

Funding

This work was supported in part by the JSPS KAKENHI Grant Number JP23H03411 and JP22K12144, and the JST Grant Number JPMJPF2017.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-94089-9.

Correspondence and requests for materials should be addressed to X.Y.