Foundation models for time series forecasting and policy evaluation in infectious disease epidemics: a modelling study

Suprabhath Kalahasti [1] **Benjamin Faucher** [1] **Boxuan Wang** [1] **Claudio Ascione** [1] **Ricardo Carbajal** [2] **Maxime Enault** [2] **Christophe Vincent Cassis** [2] **Titouan Launay** [1] **Caroline Guerrisi** [1] **Pierre-Yves Boëlle** [1] **Federico Baldo** [1] **Eugenio Valdano** [1]

Abstract

Epidemic forecasting and public health policy rely on mathematical models, but traditionally struggle in data-limited settings. We evaluated whether transformer-based foundation models can serve as a new epidemic modeling framework. We tested five models across diseases and locations, including influenza, respiratory syncytial virus (RSV), chickenpox, dengue. Foundation models demonstrated strong accuracy in short-term forecasts and predicted multiple epidemic waves. They outperformed established models on limited and irregular data. We showed foundation models can generate scenarios for policy evaluation, estimating the effect of tighter restrictions on COVID-19 cases during the Alpha variant surge in Italy in 2021. We also used them to estimate the effectiveness of the 2023 RSV immunization campaign in Paris, France. Our findings suggest foundation models can complement existing modeling approaches. Their ability to generate forecasts and counterfactual analyses with minimal data highlights their potential for public health, particularly in emergent and resource-constrained settings.

1. Introduction

Epidemic models have found applications across diseases, transmission routes and locations. Major public health agencies routinely coordinate forecasting efforts for seasonal acute respiratory illness (Biggerstaff et al., a; Fiandrino et al.) and extend to vector-borne endemic diseases, such as dengue (Ind) and West Nile Virus (Holcomb et al.). During COVID-19, models helped evaluate and design non-pharmaceutical interventions (Ruktanonchai et al.) and vaccination campaigns (Watson et al.).Despite their active development and widespread use, epidemic models have limitations in their performance and applicability. (Buckee et al.; De Angelis et al.). These data-heavy models often struggle to generalize across contexts, creating ever-growing data demands. As a result, communities with strong data infrastructures and safeguards benefit, while others may lag behind or face higher risks of personal data misuse.(Oliver et al.)

To overcome these limitations, we took a different approach of modeling epidemics using transformer-based foundation models, which have recently demonstrated strong performance on tabular and time series data (Hollmann et al., b). We investigated whether existing foundation models for time series can form the basis of a new architectural framework for epidemic modeling. We evaluated five models - TimesFM (Das et al.), Lag-Llama (based on LLaMA) (Touvron et al.; Rasul et al.), Amazon's Chronos Small (Ansari et al.; Raffel et al.), TimeGPT (Garza et al.) - and TabPFN (Hollmann et al., b; Hoo et al.; Hollmann et al., a). More specifically, we focus on predicting the incidence —ratio of infected individuals every 100,000 infected —across different diseases and pathogens.

The models were assessed on multiple epidemic tasks. Notably, none of these models were originally designed or trained for epidemic modeling or public health applications. Nevertheless, we tested their performance on epidemic prediction tasks with minimal retraining, (outperforming traditional models to good effect) exploiting the power of transformer-based models to generalize across knowledge domains.

2. Related Works

Historically, forecasting of infectious diseases largely used deterministic mathematical models (e.g : SIR) (Yang et al.; Osthus et al.) combining them with probabilistic approaches

*Equal contribution [1]Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, F75012, Paris, France [2]AP-HP, Hôpital Armand Trousseau, Paediatric Emergency Department, Paris, France. Correspondence to: Eugenio Valdano <eugenio.valdano@inserm.fr>, Suprabhath Kalahasti <suprabhath.kalahasti@iplesp.upmc.fr>.

Proceedings of the 1st ICML Workshop on Foundation Models for Structured Data, Vancouver, Canada. 2025.

for uncertainty quantification. Other approaches included the use of generalized linear models, time series models, agent-based models and metapopulation models (Shaman et al.; Chretien et al.; Nsoesie et al.). However, with advances in the availability of digital tools and technologies, diverse sources of data became accessible to researchers. This enabled the development of novel approaches to forecast infectious diseases (Ali & Cowling; Dugas et al.; Kandula et al.; Achrekar et al.; Shaghaghi et al.) harnessing multiple data streams to improve the predictive power/accuracy of the model outcomes. This further led to the use of ensemble forecasting approaches aimed at enhancing the quality of forecasts and systematically account for the uncertainties (Wu & Levinson; Sherratt et al.; McGowan et al.; Mathis et al.; Biggerstaff et al., b).

In the recent times, neural network driven approaches showed promising results in epidemic forecasting and have been quite useful in enhancing public health decision making(Panagopoulos et al.; Chen & Moraga; Liu et al.; Wang et al.). However, tradtional (statistical & mechanistic) models and deep learning models currently struggle at long term forecasting within an epidemiological context (Pre), signaling that there is scope for newer methodologies to contribute to this problem.

Recent advances in Foundation Models have shown great promise in their ability to serve a broad range of time series tasks. These models have been shown to be important, especially in showcasing their abilities in a zero-shot learning framework(Yeh et al.; Liang et al.). So far, these models have not been used in epidemic contexts. We present this work as a case study to show the utility of these models in forecasting epidemics.

3. Foundation Models for Epidemic Forecasting

We selected some of the major pre-trained foundation time-series models available, namely TabPFN-TS (time series version), Chronos, Lag-Llama, TimesFM, and TimeGPT. They have been trained using similar strategies but vary in size and architectures. TabPFN-TS frames time series forecasting a tabular regression problem to predict future time events. Unlike auto-regressive models, it is able to make multi-step ahead predictions solely using past data. TimeGPT, based on self-attention mechanism, uses an encoder-decoder architecture and is primarily trained on time-series datasets, uses rolling historical values to generate forecasts. Chronos is based on the T5 architecture, uses a tokenization approach to convert time series information into tokens. At inference, it auto-regressively samples tokens from the model and maps it back to the numerical values. TimesFM is a decoder only transformer model which tokenizes time series data into discrete patches and uses auto-regressive decoding during inference. Lag-Llama is also a deocoder only model built on Llama architecture, whose tokenization process involves using lagged features and uses Rotary Positional Encoding (RoPE) in its attention layers. It also uses an auto-regressive decoding approach during inference.

Importantly, these models can be utilized "off-the-shelf" in a zero-shot fashion, in the scope of this paper, to learn the incidence curve of infectious disease epidemics. To maximize performance of each model, we also adjusted the hyperparameters following a grid search approach and selecting the best model.

3.1. Epidemic Forecasting

We tested performance of foundation time-series models in forecasting case incidence across different diseases and pathogens - Influenza-Like-Illness (ILI), chickenpox, respiratory syncytial virus (RSV) and dengue - spanning different geographical locations and transmission routes. The models were assessed on multiple tasks, including long-term multi-season predictions (3 seasons ahead), short-term incidence forecasting (4 weeks ahead), and epidemic peak timing estimation.

3.2. Policy Evaluation

For COVID-19, we focused on the Latium region in Italy, where increasing SARS-CoV-2 incidence in February-March 2021 led to a delayed tightening of restriction measures on March 14. Using data up to February 22, 2021, we fine-tuned TabPFN-TS and estimated the impact of an earlier adoption of stricter restrictions (orange tier) by comparing the predicted incidence under this scenario with the observed trajectory, using past data from Italian regions and including the tier (yellow, orange, red) as additional covariate. The difference between the observed and counterfactual incidence provided an estimate of the number of cases that could have been averted had stricter measures been implemented earlier.

For RSV immunization, we evaluated the effect of nirsevimab introduction on bronchiolitis-related emergency room (ER) admissions among infants aged 0-12 months, stratified in three age classes: 0-3 months, 4-6 months, 7-12 months. We fine-tuned TabPFN-TS on four historical pre-immunization bronchiolitis data between 2017 and 2023 and generated age-stratified forecasts for the 2023-2024 season under an alternative scenario in which immunization had not been introduced. We measured averted ER admissions as a difference between the model prediction and the data, and estimated the age-stratified effectiveness in reducing all-cause bronchiolitis ER admission as the relative reduction in observed admissions relative to model-predicted admissions, divided by the expected nirsevimab coverage among infants in that age class.

4. Experimental Results

We evaluated the each model, namely the foundation models and the epidemiological baselines, on five datasets (ILI in France, influenza in France, RSV in France, chickenpox in France, dengue in Brazil) and across three forecasting tasks: long-term multi-season incidence forecasting, short-term incidence forecasting and peak timing forecasting. Forecasting scenarios spanned three seasons, from 2016 to 2019. For each scenario, all available data before the start of the forecasting window were used for model fitting or fine-tuning. This resulted in different epidemics having very different sizes of the training data set. For each scenario and model, we forecast the median and the 1st and 9th deciles.

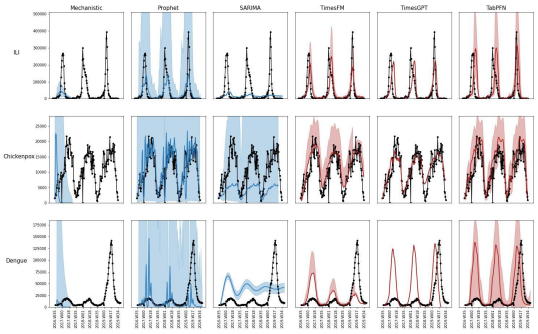

Figure 1. Long-term forecasting. Each plot compares the three epidemic seasons between 2016 and 2019 with model forecasts. Each row is a different dataset (disease) and each column is a different model. Benchmarks are in blue, foundation models are in red. The solid line is the median prediction, the shaded area encompasses the range between the 1st and the 9th decile. Forecasts start in October for ILI, Influenza, RSV and dengue, and in September for Chickenpox.

4.1. Data

For the forecasting exercise, we tested the models in five diseases/datasets - Influenza-like illness (ILI), RSV, Influenza and Chickenpox in France (Flahault et al.) ; Dengue incidence data from Brazil (Clarke et al.). For policy evaluation, we used COVID-19 incidence and tier restriction data from Italy (Badr et al.; Manica et al.). Bronchiolitis admissions to the ER in children less than 1 year-old came from seven pediatric hospitals in the region of Paris, France, over 5 seasons (October-February; years 2017-2019, 2022-2023).

4.2. Benchmark Models

Our baseline is composed of four canonical models:: two statistical models , one mechanistic model and one deep learning model. The first statistical model is Seasonal Autoregressive Integrated Moving Average (SARIMA) (Spaeder et al.) as implemented in the R package (Hyndman & Khandakar). For each scenario, the model was trained on the available data up to the last point of the training data set. The hyperparameters were calibrated using the `auto.arima` function. The second statistical model is Prophet developed by Meta (Taylor & Letham). The mechanistic model was a Susceptible-Infected-Recovered (SIR) modeled with seasonal dynamics to fit the data and generate forecasts and is similar to what was used in (Andronico et al.; Osthus et al.; Birrell et al.). The deep learning model we used was the LSTM forecasting model from Nixtla (Olivares et al., 2022; Sak et al., 2014; Elman).

4.3. Metrics

We evaluated incidence forecasts using Mean Absolute Error (MAE) divided by the dataset mean value, Mean Absolute Percentage Error (MAPE), Weighted Interval Score (WIS) (Bracher et al.) and peak timing forecasts (date at which the disease incidence peaked) using Absolute Error (AE). We then defined improvement with MAE and MAPE as the relative drop in a model's error with respect to the mechanistic model. Using the prediction interval coverage, we also assessed the calibration of the models. We considered an 80% coverage to measure calibration of the models.

4.4. Long Term Forecasting

First, we tested the performance of foundation models to forecast case incidence several seasons into the future, from October 1, 2016 to August 31, 2019, except for the mechanistic model and Chronos. This is a task where traditional modeling approaches struggle, unless periodic patterns are stable and past data are available for training over long periods. Benchmark models had a poor performance on long-term forecasts. Among them, Prophet was fairly accurate on the time series which had the longest training data - ILI and Chickenpox (see Figure 1). Its forecast, however, had a wide uncertainty. Lag-Llama was not able to perform long-term forecasts, Chronos only on chickenpox and ILI, with the limitation of being restricted to only one season. The picture was completely different for TimesFM, TimeGPT and TabPFN-TS. They showed remarkable performance on ILI, influenza and chickenpox where forecasts were both accurate and precise up to three seasons into the future, despite a small bias in the timing of the peak for ILI and influenza. For RSV and dengue, TimesFM and TabPFN-TS could forecast the timing and shape of the season, but not the peak incidence. TimeGPT could do it for dengue, not for RSV.

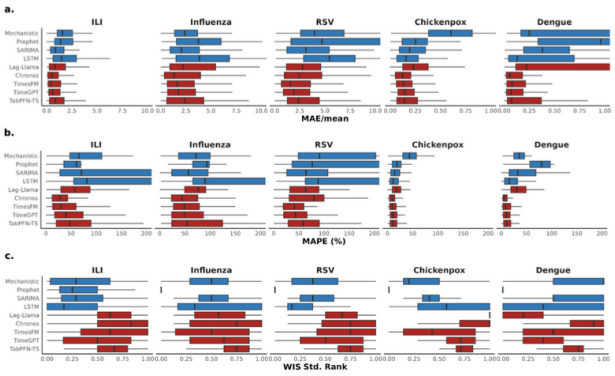

Figure 2. Model performances on short-term (4 week ahead) forecasts, across datasets. Benchmarks are in blue, foundational models are in red. (a) reports MAE divided over the dataset mean. Boxplots indicate the 2.5%, 25%, 50%, 97.5% percentiles. (b) shows the same for MAPE. (c) shows the standardized rank of WIS. WIS is a standard metric that combines accuracy and sharpness of predictive intervals. Ranks range from 0 (worst) to 1 (best), with models ranked relative to each other within each forecasting scenario.

4.5. Short Term Forecasting

We then tested the performance on shorter-term forecasts, namely predicting case incidence four weeks into the future. Each model produced nine predictions for each disease and in each of the three seasons from 2016 to 2019. Established models routinely generate short-term forecasts for both emergency response and for seasonal epidemic monitoring, but foundation models emerged here as a promising new framework to increase accuracy and reliability of those forecasts. Despite testing them on a demanding task such as predicting incidence one month into the future, foundation models showed error profiles that are comparable to, and sometimes better than, established models when long and regular training data are available. They, instead, clearly outperformed traditional approaches when data are small in size and irregular, as the example of dengue shows (see Figure 2). There, foundation models achieved very small errors averaging 10% in relative terms, corresponding to as few as weekly 2,000 cases for epidemics reaching peaks of 100,000.

4.6. Comparing TabPFN-TS with the US COVID-19 Forecast Hub

Launched in April 2020, the Forecast Hub provided real-time and retrospective forecasts of reported COVID-19 cases, COVID-19-related hospitalizations and COVID-19-related deaths at multiple spatial scales (county, state, national) from over 110 unique models. TabPFN-TS performed comparably, and in some cases better. The fact that a relatively small model with minimal computational requirements could match the performance of a state-of-the-art ensemble forecast is notable and supports incorporating foundation models.

4.7. Policy Evaluation

Our study showed that TabPFN-TS can generate counterfactual scenarios that can be used to estimate the impact of interventions - or the absence thereof - on epidemic indicators such as the incidence of cases or severe disease. The application to the tiered COVID-19 restrictions in force in Italy in 2021 illustrated this capability: focusing on a period when cases were rising in central Italy concomitantly to the spread of the Alpha variant, but restrictions were minimal, TabPFN-TS could estimate the number of COVID-19 cases that an earlier enforcement of tighter restrictions could have averted a median value of 36 detected cases (1st decile: -110, 9th decile: 182) every 100,000 inhabitants. Similarly, our evaluation of the impact of nirsevimab immunization on reducing bronchiolitis-related admissions to the ER demonstrated the potential of foundation models to rapidly evaluate the effectiveness of immunization campaigns. TabPFN-TS could predict an entire season of ER admissions while trained on limited data and provide an estimate of age-stratified effectiveness of the immunization campaign which were compatible with the results from an epidemiological study (Carbajal et al.).

5. Conclusion

We provided a broad-spectrum performance evaluation of foundation models for time-series in epidemic forecasting, showing that it has remarkable potential across different epidemic settings. Specific studies will be required to evaluate foundation models in specific public health settings to strengthen trust toward these methods. Moreover, this study includes a limited number of benchmarks, even though many models have been developed for specific epidemic contexts. Nevertheless, this study represents a starting point for a new perspective on epidemic modeling. Most importantly, integrating foundation models as a tool for public health decision-making represent a promising endeavor. Indeed, foundation models are good candidates to outperform traditional models in specific tasks, notably for long-term forecasts and for scenario analysis when limited data are available. A future research direction that merits exploring is to build foundation models specifically for epidemic tasks.

Software and Data

Code Repository: Data and Code used in this paper

Acknowledgements

This study was partially supported by Horizon Europe grant SIESTA (101131957) to Eugenio Valdano., the CIPHOD project (ANR-23-CPJ1-0212-01) to Federico Baldo., the European Union's Horizon Europe research and innovation programme Cofund SOUND.AI under the Marie Sklodowska-Curie Grant Agreement No 101081674 to Suprabhath Kalahasti. We would like to thank the annonymous reviewers for their valuable feedback/comments on the initial submission.

References

Indian Ocean temperature anomalies predict long-term global dengue trends — Science. URL https://www.science.org/doi/10.1126/science.adj4427.

On the predictability of COVID-19 - International Institute of Forecasters. URL https://forecasters.org/blog/2021/09/28/on-the-predictability-of-covid-19/.

Achrekar, H., Gandhe, A., Lazarus, R., Yu, S.-H., and Liu, B. Predicting Flu Trends using Twitter data. In *2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 702–707. doi: 10.1109/INFCOMW.2011.5928903. URL https://ieeexplore.ieee.org/document/5928903.

Ali, S. T. and Cowling, B. J. Influenza Virus: Tracking, Predicting, and Forecasting. 42: 43–57. ISSN 0163-7525, 1545-2093. doi: 10.1146/annurev-publhealth-010720-021049. URL https://www.annualreviews.org/content/journals/10.1146/annurev-publhealth-010720-021049.

Andronico, A., Paireau, J., and Cauchemez, S. Integrating information from historical data into mechanistic models for influenza forecasting. 20(10):e1012523. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1012523. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1012523.

Ansari, A. F., Stella, L., Turkmen, A. C., Zhang, X., Mercado, P., Shen, H., Shchur, O., Rangapuram, S. S., Arango, S. P., Kapoor, S., Zschiegner, J., Maddix, D. C., Wang, H., Mahoney, M. W., Torkkola, K., Wilson, A. G., Bohlke-Schneider, M., and Wang, B. Chronos: Learning the Language of Time Series. ISSN 2835-8856. URL https://openreview.net/forum?id=gerNCVqqtR.

Badr, H. S., Zaitchik, B. F., Kerr, G. H., Nguyen, N.-L. H., Chen, Y.-T., Hinson, P., Colston, J. M., Kosek, M. N., Dong, E., Du, H., Marshall, M., Nixon, K., Mohegh, A., Goldberg, D. L., Anenberg, S. C., and Gardner, L. M. Unified real-time environmental-epidemiological data for multiscale modeling of the COVID-19 pandemic. 10(1):367. ISSN 2052-4463. doi: 10.1038/s41597-023-02276-y. URL https://www.nature.com/articles/s41597-023-02276-y.

Biggerstaff, M., Alper, D., Dredze, M., Fox, S., Fung, I. C.-H., Hickmann, K. S., Lewis, B., Rosenfeld, R., Shaman, J., Tsou, M.-H., Velardi, P., Vespignani, A., Finelli, L., and for the Influenza Forecasting Contest Working Group. Results from the centers for disease control and prevention's predict the 2013–2014 Influenza Season Challenge. 16(1):357, a. ISSN 1471-2334. doi: 10.1186/s12879-016-1669-x. URL https://doi.org/10.1186/s12879-016-1669-x.

Biggerstaff, M., Johansson, M., Alper, D., Brooks, L. C., Chakraborty, P., Farrow, D. C., Hyun, S., Kandula, S., McGowan, C., Ramakrishnan, N., Rosenfeld, R., Shaman, J., Tibshirani, R., Tibshirani, R. J., Vespignani, A., Yang, W., Zhang, Q., and Reed, C. Results from the second year of a collaborative effort to forecast influenza seasons in the United States. 24:26–33, b. ISSN 1755-4365. doi: 10.1016/j.epidem.2018.02.003. URL https://www.sciencedirect.com/science/article/pii/S1755436517300889.

Birrell, P. J., Ketsetzis, G., Gay, N. J., Cooper, B. S., Presanis, A. M., Harris, R. J., Charlett, A., Zhang, X.-S., White, P. J., Pebody, R. G., and De Angelis, D. Bayesian modeling to unmask and predict influenza A/H1N1pdm dynamics in London. 108(45):18238–18243. doi: 10.1073/pnas.1103002108. URL https://www.pnas.org/doi/10.1073/pnas.1103002108.

Bracher, J., Ray, E. L., Gneiting, T., and Reich, N. G. Evaluating epidemic forecasts in an interval format. 17(2):e1008618. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1008618. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008618.

Buckee, C., Noor, A., and Sattenspiel, L. Thinking clearly about social aspects of infectious disease transmission. 595(7866):205–213. ISSN 1476-4687. doi: 10.1038/s41586-021-03694-x. URL https://www.nature.com/articles/s41586-021-03694-x.

Carbajal, R., Boelle, P.-Y., Pham, A., Chazette, Y., Schellenberger, M., Weil, C., Colas, A.-S., Lecarpentier, T., Schnuriger, A., Guedj, R., Lorrot, M., Corvol, H., and Enault, M. Real-world effectiveness of nirsevimab immunisation against bronchiolitis in infants: A case–control

study in Paris, France. 8(10):730–739. ISSN 2352-4642, 2352-4650. doi: 10.1016/S2352-4642(24)00171-8. URL https://www.thelancet.com/journals/lanchi/article/PIIS2352-4642(24)00171-8/abstract.

Chen, X. and Moraga, P. Forecasting dengue across Brazil with LSTM neural networks and SHAP-driven lagged climate and spatial effects. 25(1):973. ISSN 1471-2458. doi: 10.1186/s12889-025-22106-7. URL https://doi.org/10.1186/s12889-025-22106-7.

Chretien, J.-P., George, D., Shaman, J., Chitale, R. A., and McKenzie, F. E. Influenza Forecasting in Human Populations: A Scoping Review. 9(4):e94130. ISSN 1932-6203. doi: 10.1371/journal.pone.0094130. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0094130.

Clarke, J., Lim, A., Gupte, P., Pigott, D. M., family=Panhuis, given=W. G., p. u., and Brady, O. J. A global dataset of publicly available dengue case count data. 11(1):296. ISSN 2052-4463. doi: 10.1038/s41597-024-03120-7. URL https://www.nature.com/articles/s41597-024-03120-7.

Das, A., Kong, W., Sen, R., and Zhou, Y. A decoder-only foundation model for time-series forecasting. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 10148–10167. PMLR. URL https://proceedings.mlr.press/v235/das24c.html.

De Angelis, D., Presanis, A. M., Birrell, P. J., Tomba, G. S., and House, T. Four key challenges in infectious disease modelling using data from multiple sources. 10:83–87. ISSN 1755-4365. doi: 10.1016/j.epidem.2014.09.004. URL https://www.sciencedirect.com/science/article/pii/S175543651400053X.

Dugas, A. F., Jalalpour, M., Gel, Y., Levin, S., Torcaso, F., Igusa, T., and Rothman, R. E. Influenza Forecasting with Google Flu Trends. 8(2):e56176. ISSN 1932-6203. doi: 10.1371/journal.pone.0056176. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0056176.

Elman, J. L. Finding structure in time. 14(2):179–211. ISSN 0364-0213. doi: 10.1016/0364-0213(90)90002-E. URL https://www.sciencedirect.com/science/article/pii/036402139090002E.

Fiandrino, S., Bizzotto, A., Guzzetta, G., Merler, S., Baldo, F., Valdano, E., Urdiales, A. M., Bella, A., Celino, F., Zino, L., Rizzo, A., Li, Y., Perra, N., Gioannini, C., Milano, P., Paolotti, D., Quaggiotto, M., Rossi, L., Vismara, I., Vespignani, A., and Gozzi, N. Collaborative forecasting of influenza-like illness in Italy: The Influcast experience. 50:100819. ISSN 1755-4365. doi: 10.1016/j.epidem.2025.100819. URL https://www.sciencedirect.com/science/article/pii/S1755436525000076.

Flahault, A., Blanchon, T., Dorléans, Y., Toubiana, L., Vibert, J. F., and Valleron, A. J. Virtual surveillance of communicable diseases: A 20-year experience in France. 15(5):413–421. ISSN 0962-2802. doi: 10.1177/0962280206071639. URL https://doi.org/10.1177/0962280206071639.

Garza, A., Challu, C., and Mergenthaler-Canseco, M. TimeGPT-1. URL http://arxiv.org/abs/2310.03589.

Holcomb, K. M., Mathis, S., Staples, J. E., Fischer, M., Barker, C. M., Beard, C. B., Nett, R. J., Keyel, A. C., Marcantonio, M., Childs, M. L., Gorris, M. E., Rochlin, I., Hamins-Puértolas, M., Ray, E. L., Uelmen, J. A., DeFelice, N., Freedman, A. S., Hollingsworth, B. D., Das, P., Osthus, D., Humphreys, J. M., Nova, N., Mordecai, E. A., Cohnstaedt, L. W., Kirk, D., Kramer, L. D., Harris, M. J., Kain, M. P., Reed, E. M. X., and Johansson, M. A. Evaluation of an open forecasting challenge to assess skill of West Nile virus neuroinvasive disease prediction. 16(1):11. ISSN 1756-3305. doi: 10.1186/s13071-022-05630-y. URL https://doi.org/10.1186/s13071-022-05630-y.

Hollmann, N., Müller, S., Eggensperger, K., and Hutter, F. TabPFN: A Transformer That Solves Small Tabular Classification Problems in a Second. a. URL https://openreview.net/forum?id=eu9fVjVasr4.

Hollmann, N., Müller, S., Purucker, L., Krishnakumar, A., Körfer, M., Hoo, S. B., Schirrmeister, R. T., and Hutter, F. Accurate predictions on small data with a tabular foundation model. 637(8045):319–326, b. ISSN 1476-4687. doi: 10.1038/s41586-024-08328-6. URL https://www.nature.com/articles/s41586-024-08328-6.

Hoo, S. B., Müller, S., Salinas, D., and Hutter, F. The Tabular Foundation Model TabPFN Outperforms Specialized Time Series Forecasting Models Based on Simple Features. URL https://openreview.net/forum?id=H02X7RO3OC#discussion.

Hyndman, R. J. and Khandakar, Y. Automatic Time Series Forecasting: The forecast Package for R. 27:1–22. ISSN 1548-7660. doi: 10.18637/jss.v027.i03. URL https://doi.org/10.18637/jss.v027.i03.

Kandula, S., Hsu, D., and Shaman, J. Subregional Nowcasts of Seasonal Influenza Using Search Trends. 19(11):e7486.

doi: 10.2196/jmir.7486. URL https://www.jmir.org/2017/11/e370.

Liang, Y., Wen, H., Nie, Y., Jiang, Y., Jin, M., Song, D., Pan, S., and Wen, Q. Foundation Models for Time Series Analysis: A Tutorial and Survey. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24, pp. 6555–6565. Association for Computing Machinery. ISBN 979-8-4007-0490-1. doi: 10.1145/3637528.3671451. URL https://dl.acm.org/doi/10.1145/3637528.3671451.

Liu, Z., Wan, G., Prakash, B. A., Lau, M. S., and Jin, W. A Review of Graph Neural Networks in Epidemic Modeling. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24, pp. 6577–6587. Association for Computing Machinery. ISBN 979-8-4007-0490-1. doi: 10.1145/3637528.3671455. URL https://dl.acm.org/doi/10.1145/3637528.3671455.

Manica, M., Guzzetta, G., Riccardo, F., Valenti, A., Poletti, P., Marziano, V., Trentini, F., Andrianou, X., Mateo-Urdiales, A., family=Manso, given=Martina, p. u., Fabiani, M., Vescio, M. F., Spuri, M., Petrone, D., Bella, A., Iavicoli, S., Ajelli, M., Brusaferro, S., Pezzotti, P., and Merler, S. Impact of tiered restrictions on human activities and the epidemiology of the second wave of COVID-19 in Italy. 12(1):4570. ISSN 2041-1723. doi: 10.1038/s41467-021-24832-z. URL https://www.nature.com/articles/s41467-021-24832-z.

Mathis, S. M., Webber, A. E., León, T. M., Murray, E. L., Sun, M., White, L. A., Brooks, L. C., Green, A., Hu, A. J., Rosenfeld, R., Shemetov, D., Tibshirani, R. J., McDonald, D. J., Kandula, S., Pei, S., Yaari, R., Yamana, T. K., Shaman, J., Agarwal, P., Balusu, S., Gururajan, G., Kamarthi, H., Prakash, B. A., Raman, R., Zhao, Z., Rodríguez, A., Meiyappan, A., Omar, S., Baccam, P., Gurung, H. L., Suchoski, B. T., Stage, S. A., Ajelli, M., Kummer, A. G., Litvinova, M., Ventura, P. C., Wadsworth, S., Niemi, J., Carcelen, E., Hill, A. L., Loo, S. L., McKee, C. D., Sato, K., Smith, C., Truelove, S., Jung, S.-m., Lemaitre, J. C., Lessler, J., McAndrew, T., Ye, W., Bosse, N., Hlavacek, W. S., Lin, Y. T., Mallela, A., Gibson, G. C., Chen, Y., Lamm, S. M., Lee, J., Posner, R. G., Perofsky, A. C., Viboud, C., Clemente, L., Lu, F., Meyer, A. G., Santillana, M., Chinazzi, M., Davis, J. T., Mu, K., Pastore y Piontti, A., Vespignani, A., Xiong, X., Ben-Nun, M., Riley, P., Turtle, J., Hulme-Lowe, C., Jessa, S., Nagraj, V. P., Turner, S. D., Williams, D., Basu, A., Drake, J. M., Fox, S. J., Suez, E., Cojocaru, M. G., Thommes, E. W., Cramer, E. Y., Gerding, A., Stark, A., Ray, E. L., Reich, N. G., Shandross, L., Wattanachit, N., Wang, Y., Zorn, M. W., Aawar, M. A., Srivastava, A., Meyers, L. A., Adiga, A., Hurt, B., Kaur, G., Lewis, B. L., Marathe, M., Venkatramanan, S., Butler, P., Farabow, A., Ramakrishnan, N., Muralidhar, N., Reed, C., Biggerstaff, M., and Borchering, R. K. Evaluation of FluSight influenza forecasting in the 2021–22 and 2022–23 seasons with a new target laboratory-confirmed influenza hospitalizations. 15(1):6289. ISSN 2041-1723. doi: 10.1038/s41467-024-50601-9. URL https://www.nature.com/articles/s41467-024-50601-9.

McGowan, C. J., Biggerstaff, M., Johansson, M., Apfeldorf, K. M., Ben-Nun, M., Brooks, L., Convertino, M., Erraguntla, M., Farrow, D. C., Freeze, J., Ghosh, S., Hyun, S., Kandula, S., Lega, J., Liu, Y., Michaud, N., Morita, H., Niemi, J., Ramakrishnan, N., Ray, E. L., Reich, N. G., Riley, P., Shaman, J., Tibshirani, R., Vespignani, A., Zhang, Q., and Reed, C. Collaborative efforts to forecast seasonal influenza in the United States, 2015–2016. 9(1):683. ISSN 2045-2322. doi: 10.1038/s41598-018-36361-9. URL https://www.nature.com/articles/s41598-018-36361-9.

Nsoesie, E. O., Brownstein, J. S., Ramakrishnan, N., and Marathe, M. V. A systematic review of studies on forecasting the dynamics of influenza outbreaks. 8 (3):309–316. ISSN 1750-2659. doi: 10.1111/irv.12226. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/irv.12226.

Olivares, K. G., Challú, C., Garza, F., Canseco, M. M., and Dubrawski, A. NeuralForecast: User friendly state-of-the-art neural forecasting models. PyCon Salt Lake City, Utah, US 2022, 2022. URL https://github.com/Nixtla/neuralforecast.

Oliver, N., Lepri, B., Sterly, H., Lambiotte, R., Deletaille, S., De Nadai, M., Letouzé, E., Salah, A. A., Benjamins, R., Cattuto, C., Colizza, V., family=Cordes, given=Nicolas, p. u., Fraiberger, S. P., Koebe, T., Lehmann, S., Murillo, J., Pentland, A., Pham, P. N., Pivetta, F., Saramäki, J., Scarpino, S. V., Tizzoni, M., Verhulst, S., and Vinck, P. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. 6(23):eabc0764. doi: 10.1126/sciadv.abc0764. URL https://www.science.org/doi/10.1126/sciadv.abc0764.

Osthus, D., Hickmann, K. S., Caragea, P. C., Higdon, D., and Valle, S. Y. D. Forecasting seasonal influenza with a state-space SIR model. 11(1):202–224. ISSN 1932-6157, 1941-7330. doi: 10.1214/16-AOAS1000. URL https://projecteuclid.org/journals/annals-of-applied-statistics/volume-11/issue-1/Forecasting-seasonal-influenza-with-a-state-space
10.1214/16-AOAS1000.full.

Panagopoulos, G., Nikolentzos, G., and Vazirgiannis, M. Transfer Graph Neural Networks for Pandemic Forecasting. 35(6):4838–4845. ISSN 2374-3468. doi: 10.1609/aaai.v35i6.16616. URL https://ojs.aaai.org/index.php/AAAI/article/view/16616.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. 21(140):1–67. ISSN 1533-7928. URL http://jmlr.org/papers/v21/20-074.html.

Rasul, K., Ashok, A., Williams, A. R., Khorasani, A., Adamopoulos, G., Bhagwatkar, R., Biloš, M., Ghonia, H., Hassen, N., Schneider, A., Garg, S., Drouin, A., Chapados, N., Nevmyvaka, Y., and Rish, I. Lag-Llama: Towards Foundation Models for Time Series Forecasting. URL https://openreview.net/forum?id=jYluzCLFDM.

Ruktanonchai, N. W., Floyd, J. R., Lai, S., Ruktanonchai, C. W., Sadilek, A., Rente-Lourenco, P., Ben, X., Carioli, A., Gwinn, J., Steele, J. E., Prosper, O., Schneider, A., Oplinger, A., Eastham, P., and Tatem, A. J. Assessing the impact of coordinated COVID-19 exit strategies across Europe. 369(6510):1465–1470. doi: 10.1126/science.abc5096. URL https://www.science.org/doi/10.1126/science.abc5096.

Sak, H., Senior, A., and Beaufays, F. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In *Interspeech 2014*, pp. 338–342, 2014. doi: 10.21437/Interspeech.2014-80.

Shaghaghi, N., Calle, A., and Qian, Y. eVision: Influenza Forecasting Using CDC, WHO, and Google Trends Data. In *2020 IEEE / ITU International Conference on Artificial Intelligence for Good (AI4G)*, pp. 38–45. doi: 10.1109/AI4G50087.2020.9311072. URL https://ieeexplore.ieee.org/document/9311072.

Shaman, J., Karspeck, A., Yang, W., Tamerius, J., and Lipsitch, M. Real-time influenza forecasts during the 2012–2013 season. 4(1):2837. ISSN 2041-1723. doi: 10.1038/ncomms3837. URL https://www.nature.com/articles/ncomms3837.

Sherratt, K., Gruson, H., Grah, R., Johnson, H., Niehus, R., Prasse, B., Sandmann, F., Deuschel, J., Wolffram, D., Abbott, S., Ullrich, A., Gibson, G., Ray, E. L., Reich, N. G., Sheldon, D., Wang, Y., Wattanachit, N., Wang, L., Trnka, J., Obozinski, G., Sun, T., Thanou, D., Pottier, L., Krymova, E., Meinke, J. H., Barbarossa, M. V., Leithauser, N., Mohring, J., Schneider, J., Wlazlo, J., Fuhrmann, J., Lange, B., Rodiah, I., Baccam, P., Gurung, H., Stage, S., Suchoski, B., Budzinski, J., Walraven, R., Villanueva, I., Tucek, V., Smid, M., Zajicek, M., Perez Alvarez, C., Reina, B., Bosse, N. I., Meakin, S. R., Castro, L., Fairchild, G., Michaud, I., Osthus, D., Alaimo Di Loro, P., Maruotti, A., Eclerova, V., Kraus, A., Kraus, D., Pribylova, L., Dimitris, B., Li, M. L., Saksham, S., Dehning, J., Mohr, S., Priesemann, V., Redlarski, G., Bejar, B., Ardenghi, G., Parolini, N., Ziarelli, G., Bock, W., Heyder, S., Hotz, T., Singh, D. E., Guzman-Merino, M., Aznarte, J. L., Morina, D., Alonso, S., Alvarez, E., Lopez, D., Prats, C., Burgard, J. P., Rodloff, A., Zimmermann, T., Kuhlmann, A., Zibert, J., Pennoni, F., Divino, F., Catala, M., Lovison, G., Giudici, P., Tarantino, B., Bartolucci, F., Jona Lasinio, G., Mingione, M., Farcomeni, A., Srivastava, A., Montero-Manso, P., Adiga, A., Hurt, B., Lewis, B., Marathe, M., Porebski, P., Venkatramanan, S., Bartczuk, R. P., Dreger, F., Gambin, A., Gogolewski, K., Gruziel-Slomka, M., Krupa, B., Moszyński, A., Niedzielewski, K., Nowosielski, J., Radwan, M., Rakowski, F., Semeniuk, M., Szczurek, E., Zielinski, J., Kisielewski, J., Pabjan, B., Holger, K., Kheifetz, Y., Scholz, M., Przemyslaw, B., Bodych, M., Filinski, M., Idzikowski, R., Krueger, T., Ozanski, T., Bracher, J., and Funk, S. Predictive performance of multi-model ensemble forecasts of COVID-19 across European nations. 12:e81916. ISSN 2050-084X. doi: 10.7554/eLife.81916. URL https://doi.org/10.7554/eLife.81916.

Spaeder, M. C., Stroud, J. R., and Song, X. Time-series model to predict impact of H1N1 influenza on a children's hospital. 140(5):798–802. ISSN 1469-4409. doi: 10.1017/S0950268811001749.

Taylor, S. J. and Letham, B. Forecasting at scale. ISSN 2167-9843. URL https://peerj.com/preprints/3190.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., and Lample, G. LLaMA: Open and Efficient Foundation Language Models. URL http://arxiv.org/abs/2302.13971.

Wang, L., Adiga, A., Chen, J., Sadilek, A., Venkatramanan, S., and Marathe, M. CausalGNN: Causal-Based Graph Neural Networks for Spatio-Temporal Epidemic Forecasting. 36(11):12191–12199. ISSN 2374-3468. doi: 10.1609/aaai.v36i11.21479. URL https://ojs.aaai.org/index.php/AAAI/article/view/21479.

Watson, O. J., Barnsley, G., Toor, J., Hogan, A. B., Winskill, P., and Ghani, A. C. Global impact of the first year of COVID-19 vaccination: A mathematical

modelling study. 22(9):1293–1302. ISSN 1473-3099, 1474-4457. doi: 10.1016/S1473-3099(22)00320-6. URL https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(22)00320-6/fulltext.

Wu, H. and Levinson, D. The ensemble approach to forecasting: A review and synthesis. 132:103357. ISSN 0968-090X. doi: 10.1016/j.trc.2021.103357. URL https://www.sciencedirect.com/science/article/pii/S0968090X21003594.

Yang, W., Cowling, B. J., Lau, E. H. Y., and Shaman, J. Forecasting Influenza Epidemics in Hong Kong. 11(7):e1004383. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1004383. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004383.

Yeh, C.-C. M., Dai, X., Chen, H., Zheng, Y., Fan, Y., Der, A., Lai, V., Zhuang, Z., Wang, J., Wang, L., and Zhang, W. Toward a Foundation Model for Time Series Data. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, CIKM '23, pp. 4400–4404. Association for Computing Machinery. ISBN 979-8-4007-0124-5. doi: 10.1145/3583780.3615155. URL https://dl.acm.org/doi/10.1145/3583780.3615155.

Table 1. Data description. Summary features of the different datasets used.

DATASET	FIRST DATE	LAST DATE	POINTS	SEASONS	MEAN	MEDIAN	MAX	FREQ.
ILI	1984-10	2025-01	2092	41	60356	17426	1001824	WEEKLY
INFLUENZA	2013-12	2024-01	527	6	3275	30	73442	WEEKLY
RSV	2014-10	2024-01	370	5	2435	373	27912	WEEKLY
CHICKENPOX	1990-12	2025-01	1771	35	12101	12034	36298	WEEKLY
DENGUE	2015-01	2022-12	413	7	28574	13716	151784	WEEKLY
COVID-19	2020-11	2021-05	178	1	26	23	51	DAILY
BRONCHIOL.	2017-10	2024-02	200	5	157	112	559	FORTN.

A. Data

Réseau Sentinelles is a network of general practitioners distributed across the country who report the number of ILI cases observed in consultations on a weekly basis1. ILI cases are defined as a sudden onset of fever exceeding 39°C, with myalgia or respiratory symptoms. Each week, a subset of these cases undergo virological testing for various respiratory viruses. Test results were then used in combination with ILI data to estimate the incidence of influenza and respiratory syncytial virus (RSV) since 2013 as done in (Osthus et al.). Chickenpox incidence data were also obtained from the same system. A chickenpox case was defined as presenting the characteristic rash (erythematous-vesicular eruption lasting 3-to-4 days, pruritic, followed by a drying phase) with a sudden onset of mild fever (between 37.5 and 38 °C).

In Italy during COVID-19, the tier classification was determined automatically based on a set of epidemiological indicators, including the weekly incidence rate per 100,000 inhabitants, ICU occupancy rates, and the reproduction number (Rt), with stricter measures triggered as thresholds were exceeded(Manica et al.). Yellow tier enforced light restrictions, with limited curfews and indoor dining allowed, while orange imposed stricter measures, including travel restrictions between municipalities and the closure of bars and restaurants. The red tier had the most severe restrictions, including stay-at-home orders, school closures, and the shutdown of non-essential businesses.

In France, the bronchiolitis 2022-2023 season saw the wide scale introduction of nirsevimab, a monoclonal antibody providing passive immunization against bronchiolitis caused by RSV. Immunization was rolled out in September 2023 among children who were less than 6 months-old at this date, then at birth from October 2023 to December 2023. Coverage was assessed among children visiting the ER for conditions other than bronchiolitis(Carbajal et al.).