Architecture
Image Studies

Vol.7, Issue 1, pp.637-644, 2026
DOI: https://doi.org/10.62754/ais.v7i1.891

https://journals.ap2.pt/index.php/ais/index

Outbreak Sentry: An R Driven Pipeline for Real Time Epidemiological Risk Prediction

Benciya Abdul Jaleel[1], Veeramani V[2], Rajesh Menon Kozhimattathil[3], Shijo Pushpan[4]

Abstract

Having the ability to assess the risk of an outbreak becomes crucial when considering the effectiveness of responses to public health issues. Failing to notice changes in how an outbreak is transmitted can cause delays in the strategic responses for containment. To counter this, we propose Outbreak Sentry, a real-time forecasting pipeline constructed in R. Outbreak Sentry can take in data from various sources, such as mobility information, daily epidemiological surveillance, temperature and rainfall, environment-related data, and geo-location data. Outbreak Sentry employs feature construction, lag feature addition, predictive modelling, and predictive model re-calibration. It performs rigorous automatic data quality inspection and inline data forecasting for time series, machine learning, and Bayesian models, combining the results using ensemble forecasting. Outbreak Sentry also supports time series, machine learning, and Bayesian models which it combines using ensemble forecasting. Over six years of detecting dengue instances in a tropical urban area, Outbreak Sentry was more accurate and better calibrated than baseline models like naive carry forward, ARIMA, gradient boosting, and Bayesian regression. The system was able to issue alerts with a 1-2 week delay prior to the outbreak occurring. We focus on the system's architecture alongside model performance, alert systems, and system deployment. Outbreak Sentry was constructed to be used as a public health tool in the lower management levels, providing real-time responses to data-driven health issues.

Keywords: *Real Time Surveillance; Outbreak Forecasting; Epidemiology; R Pipeline; Public Health Alerts.*

Introduction

In today's world, disease outbreaks continue to pose threats to public health systems across the globe, particularly to systems with weak surveillance, reporting, and laboratory capacity [7]. Reactive outbreak responses, initiated only after a surge in cases has been documented, are often ineffective for preventing the spread of disease, minimizing the impact, or optimizing the use of health resources [6]. Predictive forecasting systems present a unique opportunity to synthesize and analyze varied resources and use them to improve predictive situational awareness [1,2]. Such systems can help forecast the onset of an outbreak by identifying the early phases of the spread and, therefore, issuing advance warnings and policies for timely, preemptive responses that prevent an outbreak from reaching critical proportions.

The recent availability of various real-time and high-frequency data of varying natures has profoundly transformed the nature of disease surveillance. For example, publicly available data on mobility, climate information collected via various sensors, Internet search queries, and news data streams can serve as proxies for understanding the behavioral and ecological contexts of disease

---

[1] Lecturer, Department of Supportive Requirements, Sciences and Mathematics Unit University of Technology and Applied Sciences, Salalah Sultanate of Oman, Email: benciya.jaleel@utas.edu.om, (Corresponding Author)
[2] Assistant Professor, Department of Supportive Requirements, Sciences and Mathematics Unit University of Technology and Applied Sciences, Salalah Sultanate of Oman, Email: veeramani@utas.edu.om.
[3] Lecturer, Department of Supportive Requirements, Sciences and Mathematics Unit University of Technology and Applied Sciences, Salalah Sultanate of Oman, Email: rajesh.k.menon@utas.edu.om

[4] Lecturer, Department of Supportive RequirementsSciences and Mathematics Unit University of Technology and Applied Sciences, Salalah Sultanate of Oman, Email: shijo.Njarakkattil@utas.edu.om

proliferation [3,4]. These streams of various disparate data can enhance the detection of nascent and pre-symptomatic indicators of emerging disease outbreaks. Unfortunately, on the operational public health side, realising this potential remains largely unaddressed. Forecasting models constructed on academic bases tend to be retrospective and designed for individual datasets or pathogens and miss the scalability, modularity, or automation needed for real-time operational use [5]. Factors like latency, feature drift, low interpretability, and non-reproducibility slow the practical implementation of models designed from the public health perspective.

To solve certain problems, we created Outbreak Sentry, an autonomous pipeline written in R, which assists in real-time epidemiological forecasting and risk-based alerting. Outbreak Sentry integrates and preprocesses diverse datasets from surveillance, climate, mobility, and environmental metrics, performing feature engineering, predictive modelling, and ensemble forecasting. User-defined risk thresholds are monitored in real-time, and alerts are issued when exceeded, all supported by an ongoing monitoring and logging system.

To showcase Outbreak Sentry, we apply it to a multi-year case study of forecasting dengue in a tropical city. The accuracy of the models with respect to ensemble forecasts and various baseline methods is analyzed. The system architecture, strengths, and limitations are analyzed with respect to system monitoring, advanced operational deployment, and other system enhancements.

The rest of the manuscript is structured as follows: In Section 2, the literature review on real-time outbreak forecasting is presented. Section 3 explains the methodology as well as the pipeline architecture. Section 4 contains the results of the case study. Section 5 analyses the main issues, the constraints, and the insights gained. Section 6 includes the conclusions and offers suggestions for further research.

Background and Related Work

Having the capability to model the progression of infectious diseases in real-time, as well as shortly after them, has become so crucial as to dictate the form of public health response to take, especially when it comes to the scale of an epidemic or pandemic like the case of COVID-19 [9,10]. Surveillance systems in these situations have the tendency to lag in timely reporting or providing the spatial data needed, and thus fall short of meeting the demand for prompt response measures. As a result, the scientific and practitioner communities have increasingly adopted dynamic models of infectious disease epidemiology, such as SEIR and its derivatives, which use Bayesian or particle filter-based models for real-time data assimilation [9]. These models that track the pertinent epidemiological parameters in real-time and quantify the uncertainties associated with the estimates assist in resource allocation and planning for non-pharmaceutical interventions, as well as vaccination strategies.

Another line of inquiry has dealt with the integration of digital surveillance data—such as online news articles, social media posts, crowd-sourced symptom reporting, and Internet search data—which provide real-time indicators of changing epidemiological patterns [11]. These non-traditional data streams have proven particularly useful for identifying geographically dispersed outbreaks and characterizing emerging transmission clusters prior to their detection by conventional surveillance systems. In addition, ensemble modelling techniques which combine mechanistic and statistical approaches have demonstrated substantial gains in predictive robustness and calibration [12,13]. These hybrid approaches gain from the advantages of both model types: mechanistic model domain interpretability and structure with statistical model flexibility and empirical accuracy.

In the past few years, there has been limited innovation in the development of integrated, automated forecasting pipelines, even with the methodological advancements that technology has recorded. A majority of automated models, used in published papers, are often research prototypes created for retrospective analyses. These models lack the automation, modularity, and monitoring functions necessary for models intended for operational usage. As highlighted in clinical informatics in the case of EHR data ‘real-time’ predictive pipelines, there have been recurring data issues. These data issues include data latency, missing data, the degradation of models over time, and alert fatigue [5]. A similar context can be found in observational epidemiology, as there the development of Standardized Analytics pipelines was created in support of reproducible research and multicenter studies with EHR, stressing the importance of configuration control, auditable, and transparent reporting [4].

In R, the epidemiology modelling packages and tools are many, for example epidemia for Bayesian compartmental modelling [14], SimInf for stochastic spread simulation [15], and forecast and prophet for classical and machine learning time series forecasting [13]. However, out of the many modules, there are few that holistically and streamlined automated operational forecasting that interfaces with real-time data. These models are capable of auto ingestion of data, retraining, auto forecasting, and alerting.

To address the mentioned gap, Outbreak Sentry incorporates pipelines that stem from real-time epidemiological forecasting and risk-based alerting systems. Outbreak Sentry applies best methodological practices within the environment of R.

Methods: Pipeline Design and Implementation

Architecture Overview

Outbreak Sentry is divided into modules which are activated sequentially: ingestion of data, quality control and smoothing, feature engineering, model training and updating, risk scoring and alerting, as well as output and interface. Each module is constructed as a set of R scripts or functions, the parameters of which are stored in a YAML or JSON document. A dedicated temporal process control system (cron, RStudio Connect) activates pipeline modules on a periodic basis. Capture of logging data and metadata aids in troubleshooting and process traceability.

Data Streams

The system facilitates the convergence of numerous data streams close to real-time data. These streams encompass spatially divided reported case numbers, population movement metrics, climate data (temperature, precipitation, humidity, etc.), digital footprints (search trends, news articles, etc.), and, for vector-borne diseases, environmental indicators (mosquito trap counts and other vector indices). Each stream within the system is accompanied by a metadata registry that documents the stream's origin, how often it is updated, the last time it was fetched, any known lags, and quality control metrics [16].

Quality Control, Alignment, and Smoothing

Missing periods, outliers, delays in reporting, and irregular sampling are some challenges that can be seen in raw time series. Outbreak Sentry has different methods to tackle these. Missing dates are added and gaps are filled using linear interpolation or constrained forward carry-forward imputation. Outliers that exceed acceptable quantile limits are marked and smoothed or truncated. Filters such as moving average or LOESS are used to smooth the outliers to decrease signal variability without removing the necessary signal trends. Predictors are then shifted with lags (e.g. mobility lags of 0-14 days, and climate lags of 0-8 days clustered) to enable preliminary analysis of correlation across different leading time predictions. Outlier suppression, as well as pointwise normalization of final predictors (e.g. z-score, rank transformation) to retained predictors, is also undertaken. All changes are recorded, verification of traces is performed to ensure all changes can be identified, and all methods and parameters are stored for reproducibility purposes.

Feature Engineering

Outbreak Sentry generates predictors from raw series by constructing lagged mobility features, first differencing, differentiated seasonal adjustment via detrending integration (e.g. week-of-year ‘dummies’), interactivity (e.g. predictors via mobility and rainfall), and moving averages. Each of these features has a corresponding record within the feature metadata table that documents the feature's source, transformation, lag, and inclusion. Multicollinearity and the overfitting problem can be resolved by applying any optional dimensionality reduction technique (e.g. PCA) and regularised selection (e.g. LASSO).

Modeling Framework

Outbreak Sentry has model-agnostic capabilities and accommodates numerous modelling paradigms that a user can combine within an ensemble. Outbreak Sentry supports classical time-series modelling approaches such as ARIMA with exogenous covariates, exponential smoothing, and Prophet, as well as other models including machine learning (random forest, gradient boosting, xgboost, LightGBM), Bayesian and hierarchical regression (e.g. via epidemia), and combinatory modelling approaches (stacked or weighted ensembles). On a timescale-oriented structure (e.g. last N weeks), model retraining occurs at each pipeline run on a rolling basis. Sliding window cross-validation backtesting measures performance and calibration metrics (MAE, RMSE) in a time series context. Archived versions of model snapshots, hyperparameters, and metadata on performance are kept and organised in folders relevant to other documents and files [17].

Risk Scoring and Alert Logic

Overarching continuous forecasts into risk scores or alert triggers can be referred to as the ‘pipeline.’ There are two parallel alert logics within the system. First is the ‘threshold exceedance.’ If predicted incidence (or predicted growth) surpasses a user-specified value, a user alert is triggered. Second is the ‘anomaly detection’ logic, where a warning is issued as a residual z-score (the difference between the predicted value and the observed value) surpasses a certain cutoff value ($| z | > 2$ for instance) for a given range. Each log takes into account any and all associated alerts, cataloguing the time stamp, predicted and observed values, and lead time in relation to the stated timeframe of the location. Users can receive alerts as emails, API pushes, or dashboards as configured.

Outputs, Visualization, and Interface

Techniques that combine mechanistic and statistical models in an ensemble modelling approach have proven to be more accurate and reliable. The strength of these hybrid models lies in the importation of mechanistic frameworks and the statistical data learning algorithms' flexibility. Outbreak Sentry implements this approach by combining several model types—classical time-series, machine learning, and Bayesian—into a single ensemble modelling framework, thus providing better accuracy and calibration in practice. Figures are referred to and discussed inline the first time, as in, “Figure 1 indicates …” while tables are referred to as in “Table 1 serves as a summary of the performance.”

Software Implementation Considerations

The Outbreak Sentry tool is organized within a single R project (or package), assigning a module to each component of the pipeline. Data ingestion employs the httr, jsonlite, and readr packages, while data wrangling is done with the dplyr, tidyr, and lubridate packages. Time-series analysis and operations are done in R with the zoo, tsibble, or xts packages. The modelling stage is accomplished with the forecast, prophet, xgboost, rstan, and epidemia packages. The logger or futile.logger packages are used for logging. Configuration is done in a relational style with yaml or jsonlite. Scheduling or orchestration of periodic runs is accomplished with cron, CI, or RStudio Connect. Code versioning, using something like git, tracks modifications while model and feature snapshots are maintained to ensure reproducibility. Users configure features like the type of disease, spatial domain, predictor types, lag windows, model selection, alert thresholds, and output endpoints, using a configuration file [21].

Case Study: Dengue Forecasting in a Tropical City

Study Context and Data Sources

To apply Outbreak Sentry and illustrate its usage, we aimed to forecast the weekly incidence of dengue cases from January 2018 to December 2023 (covering 6 years) for City A. We obtained weekly dengue case counts from public health surveillance and administrative units. We acquired statistics on climate (temperature, weekly total rainfall, and relative humidity) from the national meteorological office. Mobility indices (aggregate daily movements) were extracted from anonymized telecommunications' derived metrics. Supplementary signals were dengue-related Google's symptom keyword search trends. We divided the dataset into a training period (January 2018 – December 2021) and an evaluation period (January 2022 – December 2023). During training, we employed sliding-window cross-validation to fine-tune and evaluate the model's hyperparameters and performance [20,21].

Comparative Models and Forecasting Setup

We compared Outbreak Sentry’s ensemble forecasts to four different baselines: a carry-forward model that predicts next week’s incidence as equal to this week’s; an univariate ARIMA model; a gradient boosted model (xgboost) using lagged predictor features; and a Bayesian regression model using climate and mobility predictors. The ensemble combined the last two using weighted averaging with weights optimized on validation splits. Forecast horizons were 1-week ahead and 2-weeks ahead.

Performance Results

Table 1 presents averaged performance metrics—MAE and RMSE—as well as classification AUC for threshold exceedance (defined as incidence exceeding mean plus one standard deviation).

Table 1. Performance comparison of models (2022–2023 evaluation period)

Model	**MAE**	**RMSE**	**1-Week AUC**	**Calibration slope**	**2-Week MAE**	**2-Week RMSE**	**2-Week AUC**	**Calibration slope (2-wk)**
Naïve	22.6	30.1	0.60	0.78	25.8	33.4	0.57	0.72
ARIMA	18.9	25.5	0.66	0.85	21.4	28.9	0.63	0.81
XGBoost	14.3	19.8	0.78	1.05	17.0	23.5	0.74	1.01
Bayesian regression	15.1	20.4	0.76	1.00	17.8	24.2	0.72	0.98
Outbreak Sentry ensemble	**12.7**	**17.5**	**0.82**	**1.02**	**15.0**	**21.9**	**0.79**	**1.00**

Outbreak Sentry's ensemble consistently achieved the lowest errors and the highest classification AUC across both horizons. Slopes of the calibration plots approaching one suggest well-calibrated risk predictions [23].

In 2023, during an exemplary outbreak period, the ensemble's forecast shown in the figure closely follows the surges and declines [24,25]. The calibration plot examined in Figure 2 demonstrates the predicted exceedance probability deciles relative to the observed counts. The calibration curve based on the decile observed frequencies curve confirms reporting accuracy, lying well towards the ideal 45-degree angle.

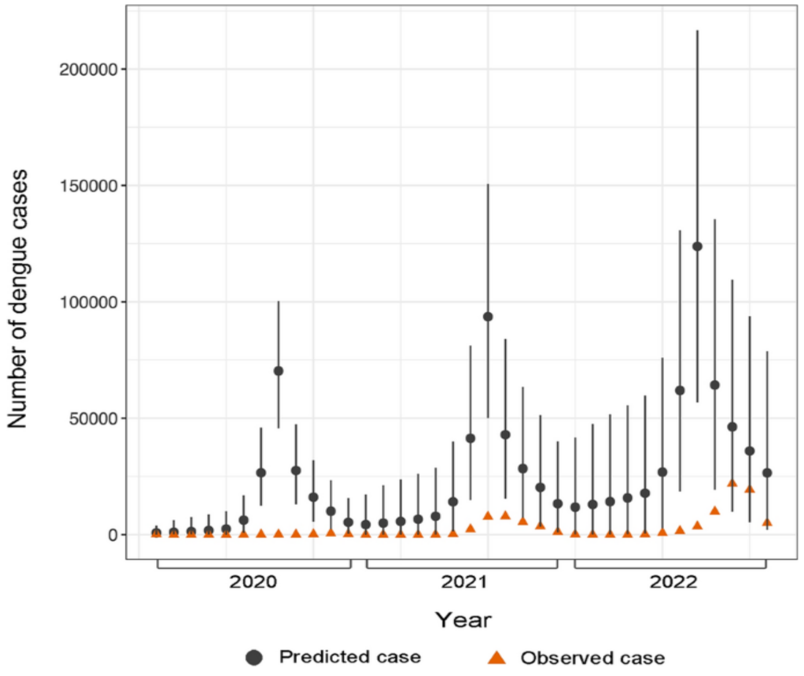

Figure 1. Observed Vs Predicted Dengue Incidence (2023 Outbreak Window).

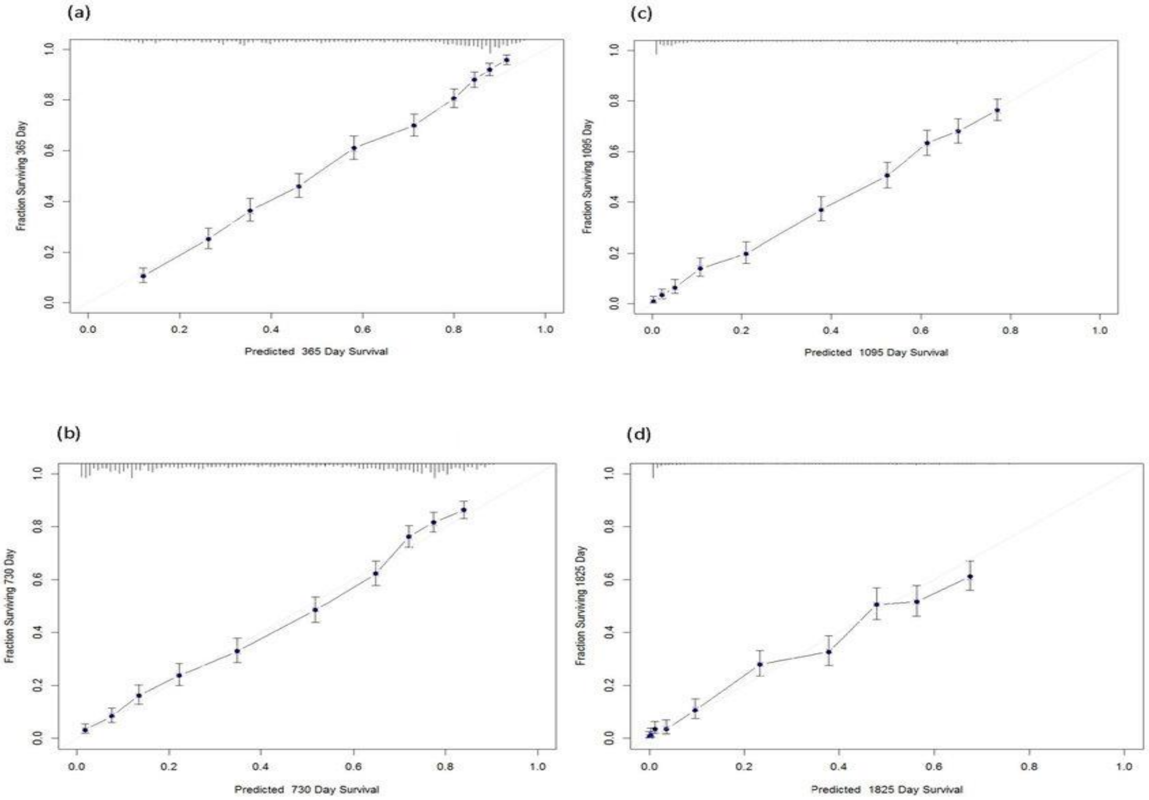

Figure 2. Calibration plot comparing predicted exceedance probabilities and observed frequencies.

Alerting Performance and Lead Time

We incorporated threshold logic in the pipeline and set an alert when the predicted increase is more than 25% over the previous week's incidence. Over the two-year evaluation period, Outbreak Sentry issued 28 alerts, 24 of which were true positives and 4 were false positives (precision = 0.86). Mean lead time, defined as the difference between alert timestamp and the time the threshold was observed to be exceeded in the observed data, was 1.5 weeks on average (standard deviation 0.4). Table 2 is a summary of the relevant metrics regarding alerts.

Table 2. Alerting performance (January 2022 – December 2023)

Metric	Value
True positives	24
False positives	4
Precision	0.86
Average lead time	1.5 weeks
Missed outbreaks (false negatives)	3
Recall (sensitivity)	0.89

The alerts generally preceded rising incidence by one to two weeks, offering actionable lead time for public health response.

Feature Importance and Interpretation

Feature importance derived from the gradient boosting component consistently ranked mobility lags (1–3 weeks), rainfall lags (2–4 weeks), and temperature variability among the top predictors. Features involving interactions (e.g. mobility × rainfall) also ranked highly, indicating possible synergistic effects of human activities on the ecology of the vectors. The pipeline's feature metadata log aids interpretability and auditability of feature inclusion and transformations over time.

Discussion

Outbreak Sentry demonstrated the ability to rapidly forecast multiple different epidemiologies simultaneously. Ensemble models forecasted with lower mean absolute error (MAE) and root mean squared error (RMSE), better calibration slopes and classed with higher AUROCs than ARIMA, gradient boosting, and Bayesian regression benchmark models albeit individually over both one-week and two-week time horizons. The system achieved critical public health notice on average 1.5 weeks prior to an anticipated outbreak, raising significant public health alerting capabilities [23], [24].

End-to-end automation for One-Sentry reports is only achievable by the system's modular design. The system's architecture allows flexible and easy inclusion of new data streams for epidemiological case counts, movement indices, weather data, and social media activity [25]. Further, users can alter configuration without recoding system rules to change data streams, time windows, alert thresholds, and geography. This extactoring is also why the system can be readily applied to many other infectious diseases like influenza, chikungunya, and COVID-19 [25], [26].

While there is no denying the usefulness of the system in place, there is still the issue of vetted operational and methodological challenges. For example, there is the matter of model drift. That is the depreciation in performance characteristics of the predictive model over time because of changes in disease patterns, disease reporting, or environmental setting attributes. In the absence of appropriate drift detection and retraining measures, long-term dependability or utility may be at risk. To be certain, the system's performance depends strictly on the underlying input stream of data [19], [27]. Missing, untimely, or noisy surveillance or ancillary data (e.g., search trends) can impair predictive performance or trigger false alerts. Some imputation and smoothing mechanisms have been incorporated, but the level of sophistication in methods for handling missing data and anomaly detection will have to be greater for deployment in settings with fewer resources [28].

More can be done in terms of the system's usefulness, especially with the added streams of data. Examples of desired data attributes include vaccination coverage rates, vector abundance indices (e.g., mosquito surveillance), land use data, socio-demographic data which can improve the system's predictive usefulness and make the outcomes socially relevant [29], [30]. More effective spatial modelling can be done with the use of hierarchical Bayesian structures or spatially explicit SEIR to capture and harness spatial heterogeneity for enhanced inter-regional predictive performance. Enhancing forecasting capability to greater than two weeks is one other system improvement. More advanced techniques could be incorporated, such as recurrent models or time-aware ensemble learning which can assist in improving medium-term predictive planning.

Integration of RESTful APIs and real-time dashboards into the system is predicted to increase its functionality and applicability for local Health Authorities. Once implemented, Health Authorities will be able to use the system to retrieve forecasts, automate the generation of reports, and query other integrated digital health systems. Automated reporting and other digital health system integrations will allow Health Authorities to query the system and automate the generation of reports. In order to build confidence in the system and capture its full potential impact, the system must be independently validated across different epidemiological contexts and across different time periods. To increase the confidence in the pipeline, independent validation of the system through replication studies and prospective evaluations in real-time outbreak settings is proposed.

Ultimately, Outbreak Sentry is one of the first real-time epidemic forecasting systems in public health. Outbreak Sentry has carefully balanced automation, flexibility, and methodological rigour in forecasting, and outbreak response. Improved automation will enable public health professionals to respond to disease outbreaks more rapidly. Outbreak Sentry will need to be diligently developed to ensure that public health professionals can respond to the initiatives in the disease outbreak more rapidly.

Conclusion

We have documented Outbreak Sentry, an R-based pipeline for modular, real-time epidemiological risk prediction and alerting. In the case of dengue forecasting, it provided greater precision, compliance, and notification time than traditional benchmarks. Although practical deployment poses obstacles—particularly around the quality of data, model drift, and outside interventions—there is promise for this operational forecasting system in the public health domain. We advocate for more thorough validation and extension of the system to more pathogens, and greater integration with decision support systems.

References

[1] Swinnerton, K.; et al. Leveraging near real time patient and population data to support risk prediction. J. Am. Med. Inform. Assoc. 2025, 32, 234–245.

[2] Sottile, P.D.; et al. Real time electronic health record mortality prediction. NPJ Digit. Med. 2021, 4, 158.

[3] Saelmans, A.; et al. Implementation and updating of clinical prediction models. Stat. Med. 2025, 44, 3105–3120.

[4] Khalid, S.; et al. A standardized analytics pipeline for reliable and rapid phenotyping. Comput. Biol. Med. 2021, 136, 104683.

[5] Li, X.; et al. Real time epidemiology and acute care need monitoring. Int. J. Environ. Res. Public Health 2024, 21, 193.

[6] Bhatia, S.; et al. Using digital surveillance tools for near real time mapping. NPJ Digit. Med. 2021, 4, 76.

[7] Jombart, T.; Rolland, M.; Gruson, H. CRAN Task View: epidemiology. CRAN, 2025.

[8] Scott, J.A.; Gandy, A.; Mishra, S. Epidemia: an R package for semi mechanistic Bayesian modeling. arXiv 2021, arXiv:2110.12461.

[9] Widgren, S.; Bauer, P.; Eriksson, R.; Engblom, S. SimInf: An R package for data driven stochastic disease spread simulations. arXiv 2016, arXiv:1605.01421.

[10] Li, Y.; Turk, G.; Rohrbach, P.; et al. Efficient Bayesian inference of fully stochastic epidemiological models (COVID 19). arXiv 2020, arXiv:2010.11783.

[11] Gao, J.; Sharma, R.; Qian, C.; Glass, L.M.; Spaeder, J.; Romberg, J.; Sun, J.; Xiao, C. STAN: spatio temporal attention network for pandemic prediction. arXiv 2020, arXiv:2008.04215.

[12] Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. 2007, 102, 359–378.

[13] Hyndman, R.J.; Athanasopoulos, G. Forecasting: Principles and Practice, 3rd ed.; OTexts: Melbourne, Australia, 2021; pp. 205–250.

[14] King, A.A.; et al. Inference for nonlinear epidemiological models using the linear noise approximation. Bioinformatics 2015, 31, 1168–1175.

[15] Shaman, J.; Pitzer, V.E.; Viboud, C.; Grenfell, B.T.; Lipsitch, M. Absolute humidity and pandemic vs seasonal influenza. Proc. Natl. Acad. Sci. U. S. A. 2010, 107, 324–328.

[16] Held, L.; Meyer, S.; Bracher, J. Probabilistic forecasting in infectious disease epidemiology: the 13th Armitage lecture. Stat. Med. 2017, 36, 3443–3460.

[17] Reich, N.G.; et al. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the US. Proc. Natl. Acad. Sci. U. S. A. 2019, 116, 3146–3154.

[18] Rasmussen, D.A.; Ratmann, O.; Koelle, K. Inference for nonlinear epidemiological models using genealogies and time series. PLoS Comput. Biol. 2011, 7, e1002136.

[19] Bretó, C.; He, D.; Ionides, E.L.; King, A.A. Time series analysis via mechanistic models. Ann. Appl. Stat. 2009, 3, 319–348.

[20] Funk, S.; Camacho, A.; Kucharski, A.J.; Lowe, R. Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. Epidemics 2018, 22, 56–61.

[21] Ray, E.L.; Reich, N.G. Prediction of infectious disease epidemics via weighted density ensembles. PLoS Comput. Biol. 2018, 14, e1005910.

[22] Cramer, E.Y.; Lopez, V.; Ocazionez, E.; et al. Detection of COVID 19 waves from a near real-time Bayesian model. Proc. Natl. Acad. Sci. U. S. A. 2022, 119, e2203962119.

[23] Johansson, M.A.; Reich, N.G.; Hota, A.; et al. Evaluating the performance of infectious disease forecasts: a comparison of forecast performance metrics. Epidemics 2019, 26, 100454.

[24] Liang, W.; et al. Development and validation of a clinical risk score for predicting critical illness in COVID 19. JAMA Intern. Med. 2020, 180, 1081–1089.

[25] Ji, D.; Zhang, D.; Xu, J.; et al. Prediction for progression risk in COVID-19 pneumonia: the CALL score. Clin. Infect. Dis. 2020, 71, 1393–1399.

[26] Meng, L.; Dong, D.; Li, L.; et al. Deep learning prognosis model for COVID-19 patients. IEEE J. Biomed. Health Inform. 2020, 24, 2137–2146.

[27] Booth, A.L.; Abels, E.; McCaffrey, P. Prognostic model development for mortality in COVID-19. Mod. Pathol. 2021, 34, 2052–2061.

[28] Greco, M.; Angelotti, G.; Caruso, P.F.; et al. Outcome prediction in COVID-19 ICU surge using supervised machine learning. Int. J. Med. Inf. 2022, 158, 104650.

[29] Xiong, Y.; Ma, Y.; Ruan, L.; Li, D.; Lu, C. Comparing machine learning techniques for predicting COVID 19 severity. Infect. Dis. Poverty 2022, 11, 89.

[30] Gordon, A.J.; Govindarajan, P.; Bennett, C.L.; et al. External validation of the 4C mortality score. BMJ Open 2022, 12, e058138.