Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

A MULTI-TEAM MULTI-MODEL COLLABORATIVE COVID-19 FORECASTING HUB FOR INDIA

Aniruddha Adiga
Benjamin Hurt
Gursharn Kaur
Bryan Lewis
Madhav Marathe
Przemyslaw Porebski
Srinivasan Venkatramanan

University of Virginia
Charlottesville, VA 22903 USA

Ambedkar Dukkipati
Tony Gracious
Shubham Gupta
Nihesh Rathod
Rajesh Sundaresan
Sarath Yasodharan

Indian Institute of Science
Bengaluru 560012, INDIA

Kantha Rao Bhimala
Vidyadhar Mudkavi
Gopal Krishna Patra

CSIR-Fourth Paradigm Institute
Bengaluru 560037, INDIA

Siva Athreya

International Centre for Theoretical Sciences-TIFR
Bengaluru 560089, INDIA

ABSTRACT

During the COVID-19 pandemic, India has seen some of the highest number of cases and deaths. Quality of data, continuously changing policy, and public health response made forecasting extremely difficult. Given the challenges in real-time forecasting, several countries had started a multi-team collaborative effort. Inspired by these works, academic partners from India and the United States setup a repository for aggregating India-specific forecasts from multiple teams. In this paper, we describe the effort and the challenges in setting up the repository. We discuss the development of simulations of compartmental models to model specific waves of the pandemic and show that the simulation model designed specifically for the Omicron wave was able to predict the onset and peak sizes accurately. We employed a median-based ensemble model to aggregate the individual forecasts. We observed that median-based ensemble was relatively stable compared to the constituent models and was one of better performing models.

1 INTRODUCTION

The COVID-19 pandemic severely strained public healthcare systems in many countries with devastating consequences. Forecasting of cases, deaths, hospitalizations, etc., especially over the short term (a few weeks) can be highly beneficial as they help public health agencies assess and allocate medical resources (Lutz et al. 2019). India reported nearly 45 million cases out of approximately 675 million cases reported worldwide. The *alpha* wave was handled with strong non-pharmaceutical interventions, and consequently the need for sound forecasts was not evident. However, when the *Delta* variant arrived India in April 2021, there was relatively no warning from the modeling community. This coupled with its greater severity of illness compared to *alpha* had a crippling impact on the health system. At that time, there was a strong

979-8-3503-6966-3/23/$31.00 ©2023 IEEE

need to understand the spreading dynamics of COVID-19 (under antibody waning, vaccination effects, variant dynamics, etc.) and several modeling efforts were renewed. Inspired by the US and European Forecast Hubs, and Open Science practices (Bezjak et al. 2018), multiple researchers from Indian and the US academic institutions came together in August 2021 to start an India-focused Hub. The aim of this hub was (*i*) to provide a common platform for modeling and forecasting teams to contribute and store short-term COVID-19 incident cases forecasts; (*ii*) to develop a suite of models that can be analyzed and modified collaboratively, (*iii*) to stipulate common formatting for ground truth and forecast data that enables easier evaluation and reproducibility, and (*iv*) to provide interactive visualization engines for effective communication of forecasts.

From a modeling stand-point, forecasting is highly challenging during an ongoing pandemic of a novel pathogen because of (*i*) the interplay between the target-of-interest and the human behavior/pathogen dynamics; (*ii*) data inconsistencies due to heterogeneous and evolving reporting schemes; and (*iii*) lack of historical data, to name a few. Also, from an operational standpoint, several challenges arise: (*i*) the forecast processing systems operate on a strict deadline and models have to be built, tested, modified and deployed within the stipulated time; (*ii*) systems and models need to evolve continuously based on the problem, requirements, and the situation at hand; (*iii*) models need to be flexible to incorporate various datasets and be robust to partial loss of data; (*iv*) concurrent with these changes there needs to be a platform that can be automatically or quickly updated for communicating results and analyses.

Given the challenges in real-time forecasting, at the start of the pandemic, several countries had started multi-team collaborative efforts that aggregated, analyzed, and communicated the forecasts in a timely manner. Notable examples include the United States (US) COVID-19 ForecastHub (Cramer et al. 2022) (https://covid19forecasthub.org/) and the European COVID-19 Forecast Hub (https://covid19forecasthub.eu/). The Forecast Hub provides a framework for multiple teams having different models to provide their forecasts. This adds robustness by not relying too heavily on any single model or team. The forecasts from various models are aggregated using an ensemble model. A collaborative effort to build such a platform was inspired by prior frameworks deployed for forecasting such as for influenza (Reich et al. 2019), Ebola (Viboud et al. 2018), and dengue fever (Johansson et al. 2019; Yamana et al. 2016) outbreaks. A key insight obtained from these efforts was that ensemble approaches produce superior performance when compared to individual models. The collaborative efforts in the US and Europe have largely been the initiative of the public health agencies such as the Center for Disease Control and Prevention (CDC). The CDC directly interacts with modelers and uses aggregate forecasts for guiding policy making (Lutz et al. 2019) (Doms et al. 2018).

Despite the progress across the world, there were several challenges to implement the Forecast Hub in India. These include the following.

- Data availability and quality: During the course of the pandemic, data processing and standardization was a voluntary effort (https://www.covid19india.org). The reporting and granularity of data was heterogeneous across states resulting in noisy data.
- Evolving public response: The policies and responses of decision makers evolved continuously. In addition, testing capacity was limited thus making the reported cases a poor proxy for the actual infections in the community.
- Disconnect between modelers and policy makers: The partnership between the modeling consortium and the policy makers in India was not well established.

Even though such a framework was not available, its needs was realized during the course of the pandemic. Several researchers called out the need for a collaboration through opinion articles (Ganesan and Subramani 2021b). A few members of our team provided (and continue to provide) modeling expertise to the state of Karnataka's ($\sim$70 million population and $\sim$200,000 sq. km.) COVID-19 Technical Advisory Committee (TAC) . The forecasts and modeling approaches developed during the course of this work directly helped the TAC in policy making (see TAC Chairman's appreciation email for the support https://tinyurl.com/255b9msf).

Table 1: A summary of the collaborative effort.

India COVID-19 Case Forecasting Collaboration	
No. of teams	4
No. of models	10
Target	Incident Cases
Forecast horizon	$1-4$ week ahead
Forecast period	Oct. 2021 $-$ Aug. 2022 (44 weeks)
Forecast type	Probabilistic/Point
Spatial Resolution	National/state (27 locations)
Model types	Statistical, simulations of mean field ODEs, and deep learning
Ensemble type	Median

2 RESULTS

We created an end-to-end forecasting pipeline in which modelers were provided raw data in a standardized and machine-readable format. Based on this data, each modeling team then uploaded their predictions to a repository. The platform then used an ensemble model to aggregate the forecasts of the individual teams. Lastly, the forecast hub had an inbuilt visualization tool that enabled the user to observe individual and ensemble model predictions. The visualization tool also allowed the user to assess the performance of models across time.

The ensembler combines a diverse set of machine learning ideas including statistical, deep learning and mechanistic models. Models were kept relatively simple so that the resulting predictions were explainable. In addition to the implementation of the best practices from earlier efforts in the US and Europe, the pipeline also addressed specific challenges posed by the issues arising in India. We now summarize the key contributions of the effort of the Forecast Hub.

- To the best of our knowledge, this is the only such initiative where multiple teams could contribute their predictions towards the Forecast Hub for India. The basic platform is generalizable and can be used for other communicable diseases as well, providing a template for future efforts.
- Many modeling teams regularly recalibrated their models, thus providing better predictions for specific outbreaks. These methods and their implementation have been documented at the Forecast Hub and carry valuable information for future deployments.
- We demonstrate the efficacy of the simulation-based ODE model in forecasting the onset of the Omicron wave across different states of India.

In this paper, in addition to the framework, we conducted a systematic evaluation of performances of the models in the Forecast Hub. We also introduce a pandemic phase-based model evaluation technique to understand the variable performances of model forecasts during the different phases of the pandemic. Evaluation results indicate that a median-based ensemble model provides robust performance across different phases of the pandemic, even in the presence of variable set of models with varying performances.

3 METHODS: TEAMS, MODELS AND WORKFLOW

The main goal of the Forecast Hub was to provide robust short-term ($1-4$ weeks ahead) predictions for different locations (state- and national-level). The target of interest was the incident cases (new cases) $\{y_{l,t}\}_{t=1}^{T}$ corresponding to a location l until week T. The forecasting problem involves predicting forward $S-$steps ahead $\{y_{l,t}\}_{t=T+1}^{T+S}$ and we attempt to predict it by employing a variety of methods. Each location was trained independently and we drop the term l in further discussions.

In this collaborative effort, teams from different academic institutions provided forecasts for incident cases in India at multiple spatial resolutions, see Table 1 for a summary. There were no restrictions placed on the type of model employed or a team's level of expertise when joining the effort. Due to data availability and quality issues mentioned earlier, most models employed simple and light-weight frameworks. This further enabled easier interpretability of the model results while simultaneously allowing for faster implementation and modifications. The class of models employed by the teams are as follows:

Simulating Compartmental Models Teams considered the most common ordinary differential equation (ODE) compartmental model (aka mean field models) used to model the progression of infectious disease, the SIR model (Adiga et al. 2020). The SIR model partitions individuals in an N size population into three disjoint compartments: susceptible (S), infected (I) and removed or recovered (R). Assuming that everyone is Susceptible in the beginning and homogeneous mixing, transition of individuals from S to I to R is governed by a set of ODEs. the model parameters specify the transition rates from susceptible state to infected state, and then to recovered state. At time t, the transitions between population-normalized states, $s_t = S_t/N$, $i_t = I_t/N$ and $r_t = R_t/N$ are given as follows:

$$\frac{ds}{dt} = -\beta_t si, \qquad \frac{di}{dt} = \beta_t si - \gamma_t i, \qquad \frac{dr}{dt} = \gamma_t i, \tag{1}$$

where transition rates, β_t is referred to as the transmission rate ($S \to I$) and γ_t as the recovery rate $(I \to R)$. Note that transition rates are time varying to account for the varying pathogen strains, social distancing norms, etc. the transition rates can vary across time. Model training involves learning the transition rates by calibrating the model on the case time series. Note that in our case each state was trained independently. SEIR model consisting of an additional Exposed (E) compartment is another version of the compartmental models.

Auto-regressive Models This class of linear methods model the signal to be forecast using its lagged versions. In addition, J exogenous time series $\{x_j(t)\}_{j=0}^{J-1}$ and its lagged versions can be included in the model. The forecast is obtained as follows:

$$y_{t+s} = \sum_{p=0}^{P} a_{p,t}^{(s)} y_{t-p} + \sum_{j=0}^{J-1} \sum_{q=0}^{Q} b_{j,q,t}^{(s)} x_{j,t-q} + \sum_{r=1}^{R} c_{r,t}^{(s)} \varepsilon_{t-r} + \varepsilon_t \quad s = 1,2,3,\&\ 4 \tag{2}$$

where ε_t is error term coefficients, $c_{r,t}^{(s)}$ are the moving average and P is the length of the training window. Specifically, (2) is the Auto-regressive integrated moving-average model with exogenous variables model (ARIMAX).

Long Short-term Models This is a popular class of deep learning models specifically designed for modeling sequential data and have been shown to be effective in capturing long-term dependencies in the data (Hochreiter and Schmidhuber 1997). An LSTM model consists of k-stacked LSTM layers and each layer consists of T cells corresponding to input sequence length T. The output of the kth LSTM layer is fed into a fully connected layer to make the final prediction.

3.1 Teams

Teams used variants of the above mentioned models and are as follows.

3.1.1 Team: IISc-ISI

This team employed variants of the simulation model and are as follows:

SIR-IISc-ISI This model is an SIR model as described in (1). The parameters of the ODE are learnt to provide the best least-square fit between the model output and the incident cases data. The training window size is automatically optimized for the best fit.

Log-Linear-IISc-ISI This model is same as SIR-IISc-ISI but considers the logarithm of the number of cases for training of the models.

Omicron-IISc-ISI This is an SEIR model, a variant of the SIR model with an additional exposed compartment ((individuals that have been infected but are not yet infectious)) designed specifically to capture the Omicron variant. A key parameter, transmission rate, was calibrated using South Africa specific cases data. Specifically, by comparing the contact rates between the two countries during the Delta and Omicron waves. The transmission advantage thus computed is applied to Indian states, and projections for Omicron variant generated under the assumption of 60% overall susceptibility in the population.

SUTRA This model is a re-implementation of the SUTRA model (Agrawal et al. 2021) by team IISc-ISI.

3.1.2 Team: UVA

Analogues and ARIMA Two variants of the ARIMAX model described in (2) was used. Analogues uses population normalized lagged case time series from other states as exogenous variables $x_{j,t}$ and assumes the moving-average terms $c_{r,t}^{(s)} = 0, \forall r, t$ in (2). The ARIMA model does not incorporate exogenous variables, i.e., $b_{j,q,t} = 0, \forall j, q, t$ in (2). In both the models, the signal is log-transformed. Due to non-stationarity, model training is done over short segments that are assumed to be relatively stationary.

lstm The model was implemented as one LSTM layer with hidden layer of size 32, one dense layer with hidden layer of size 16, a rectified linear unit activation function, and one dropout layer (dropout rate of 0.2). The output layer is a dense layer with linear activation and L2 kernel regularization (0.01 penalty factor). The historical window size is 3 weeks. A mean squared error (MSE) loss function was used and the model was trained with the Adam optimizer with a batch size of 32. Probabilistic forecasts were generated using MCDropout. A single model was trained across all states to avoid over-fitting.

SEIR As the model name indicates this is an SEIR model. A simulation optimization approach is employed to sequentially estimate the time-varying transmissibility parameter with appropriate delays and scaling applied from simulated infections to confirmed cases. For each time step, the parameter estimates are obtained using Golden Section Search. The smoothed version of the daily varying transmissibility that best fits the ground truth is used for short-term forecasting.

3.1.3 Team: CSIR-4PI

LSTM_CSIR This team used uni-variate (LSTM_CSIR) and multi-variate LSTM (Ensemble LSTM (CSIR)) time series models with weather data and COVID-19 confirmed case data, based on observed correlations with specific humidity (positive), maximum temperature (negative), and minimum temperature (positive). The model observed specific humidity and temperature playing significant role in model enhancement across various regions in India.

3.1.4 Team: CoviHawkes

Hawkes-LSTM The team used an LSTM model for transmissibility of COVID-19 using past mobility data. Training and validation were done across various forecasting windows (7, 14, and 28 days) using case counts and mobility data.

3.2 The Weekly Workflow

A schematic of the workflow is shown in Figure 1. A Github repository was setup to facilitate a common platform for storing and accessing data across teams. The incident cases data was also processed by the administrators of the repository. It is extracted from https://data.incovid19.org/ and uploaded onto the repository for the teams to access. The teams were free to chose any source data for incident cases and some models used data published by https://github.com/CSSEGISandData/COVID-19. The teams were required to submit forecasts every Monday on or before 9.00 PM Indian Standard Time. Given the small number of teams, the administrators would send an email reminder in case teams did not submit forecasts before the stipulated time. Once the forecasts were obtained, the median and the mean ensemble forecasts

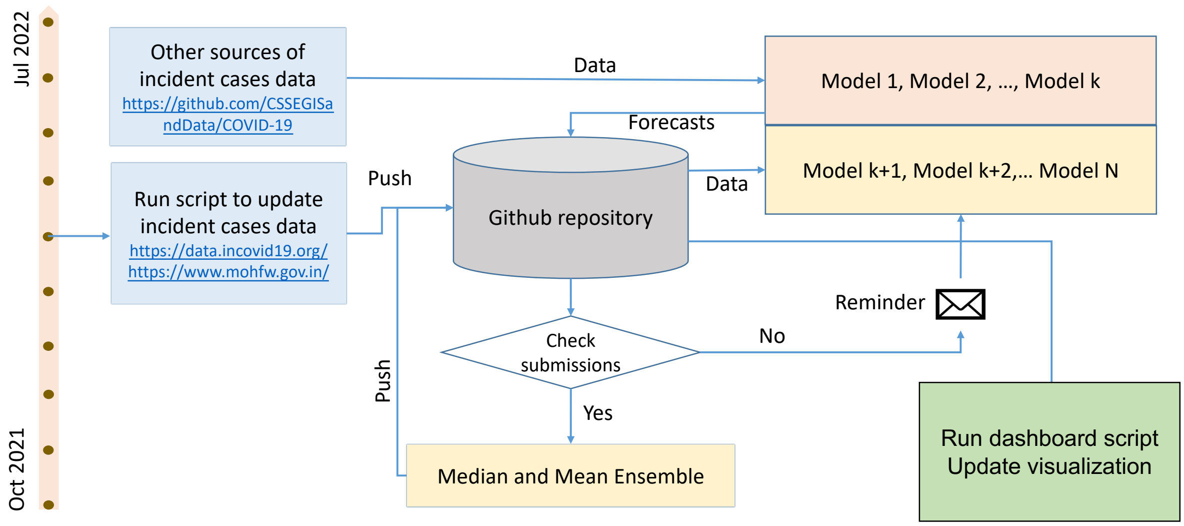


Figure 1: The workflow of the weekly processing of the incident cases data, the forecasts submissions from multiple teams, and the generation of visualizations.

were processed and uploaded to the repository. The process would culminate with the running of the dashboard scripts for setting up the visualizations (https://www.isibang.ac.in/~incovid19/dash.php).

3.2.1 Forecast File Format

A standard forecast submission format was implemented to enable faster sanity checks, ensemble model forecast generation, and dashboard updates. Although probabilistic forecasts were preferred, forecasters were allowed to provide *point* forecasts as well. Teams were requested to provide a comma separated value (csv) files with the following set of columns being mandatory:

- **avl_date**: Date (Sunday) of latest available weekly incident case data.
- **fct_date**: Date (Sunday) indicating the week for which the forecasts were made.
- **horizon**: Integer values indicating the forecast horizon (1 to 4 week ahead).
- **location**: Name of the location for which the forecasts are produced (national or one of the states).
- **value**: If probabilistic forecasts, this corresponds to the median value of the distribution, else it is the point estimate.
- **fct_lb**: If probabilistic forecasts, this corresponds to the lower bound value of the 95% confidence interval distribution, else it is assigned `NaN` for forecast submissions.
- **fct_ub**: If probabilistic forecasts, this corresponds to the upper bound value of the 95% confidence interval (CI) distribution, else it is assigned `NaN` for forecast submissions.
- **method**: Indicates the team name and the model. There were no guidelines for team names.

Note that unlike US/Europe COVID-19 ForecastHub, which expects probabilistic forecasts submitted terms of seven quantiles, we limited it to two quantiles 0.025 and 0.975 only (95% CI).

3.2.2 The Ensemble Model

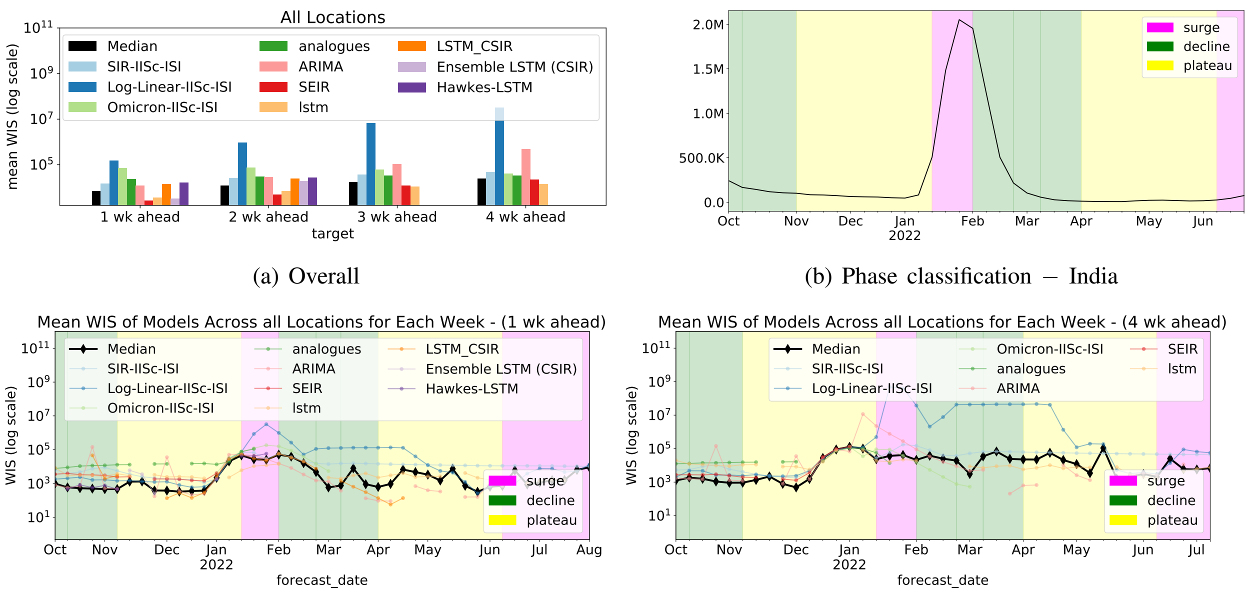
While there are several approaches proposed in the literature for aggregation (Howerton et al. 2023), we consider a median-based ensemble model to aggregate the probabilistic forecasts given by different models. Suppose the forecast from each model is given in the terms of k quantiles, each at levels $\alpha_1, \ldots, \alpha_k$. Then the $\alpha_1, \ldots, \alpha_k$ quantiles for the median-based ensemble is given by the median of α_i level quantiles from different models. In our case, we get median-based ensemble quantiles using 3 quantiles, namely **fct_lb**, **value**, and **fct_ub** across the models. This approach is useful specially when the predictive distributions from the individual models are not fully known. Unlike the quantile-averaging ensemble, median-based ensemble is less influenced by the outlier predictions. A similar model has been employed in the US COVID-19 Forecast Hub (Cramer et al. 2022) and it is observed that the forecasts from the median-based ensemble model are robust compared to the individual models. Some teams only provided median estimates in which case we assumed that the model had no uncertainty in its predictions and assumed **fct_lb**=**fct_ub**=**value**.

4 EVALUATION

We evaluate the probabilistic forecasts using the Weighted Interval Score (WIS) which is a generalization of the mean absolute error. The WIS is a proper score (Bracher et al. 2021); a lower score implies smaller error and therefore better performance. This metric is becoming popular across the epidemiological forecasting community and is being used as the primary metric of probabilistic forecast evaluation (Cramer et al. 2022). The WIS score balances different *costs* for a given forecast: (*i*) the width of the confidence interval, (*ii*) the chance of falling outside the confidence interval, which considers the distance beyond the respective lower or upper bound. For the weekly submissions of the probabilistic forecasts the teams provided the median (m), lower (.025) (l), and upper (0.975) (u) quantiles. Let $\alpha = 0.05$, corresponding to a central prediction interval of 95% CI. Given a probabilistic forecast F defined by (l, m, u), then $WIS(F, y) = \frac{1}{1+0.5}\left(\frac{1}{2}|y-m| + \frac{\alpha}{2}(u-l) + (l-y)\mathbb{1}_{y<l} + (y-u)\mathbb{1}_{y>u}\right)$, where y is the observed value and $\mathbb{1}$ is the indicator function. For the general definition of WIS computed for forecasts characterized by more than two quantiles, we refer the readers to (Bracher et al. 2021).

Figure 2 provides the performance of individual models, including the Median ensemble, across all the 27 locations. Figure 2a shows the mean WIS computed across all the forecast weeks and locations for $1-4$ week ahead forecasts and Figures 2c and 2d shows the weekly mean performance of models computed across all locations. Overall, we observe that the median-based ensemble has a performance close to the best performing model, despite variability in individual model performances across forecast weeks. Since longer horizons are harder to forecast, we observe the error increasing with the increase in horizon. But across different horizons, the ensemble model is able to provide robust forecasts. During the course of the pandemic one of challenges has been the detection of the onset of a wave and estimating peak sizes. Most forecasting systems have performed poorly during the critical phases primarily due to lack of understanding of the disease dynamics (see (Ray et al. 2021)). Hence, it is important to improve the accuracy of the forecasts in these critical phases to assess their utility.

Despite heterogeneity in the COVID-19 time series, we broadly observe three distinct phases that can be characterized by the rate of change of case counts: Surge (period of steep growth in cases), Decline (period of sharp decline in cases), and Plateau (relative stable and low case counts). Phases are subjective and several definitions exist (https://www.cdc.gov/flu/pandemic-resources/planning-preparedness/global-planning-508.html). In our case, the classification is mainly for analyzing the relative performance of models during different phases. We use the algorithm in Adiga et al. (2022) for mapping each time point into a particular phase. We provide an example of the phases in Figure 2b where each week is color coded with the corresponding phase. During the time of deployment of our systems (Oct 2021) the case counts across India were declining and plateaued during the subsequent months. In order to understand the behavior of models during different phases, each week in Figures 2c and 2d is also color coded with respective phase.



(a) Overall

(b) Phase classification — India

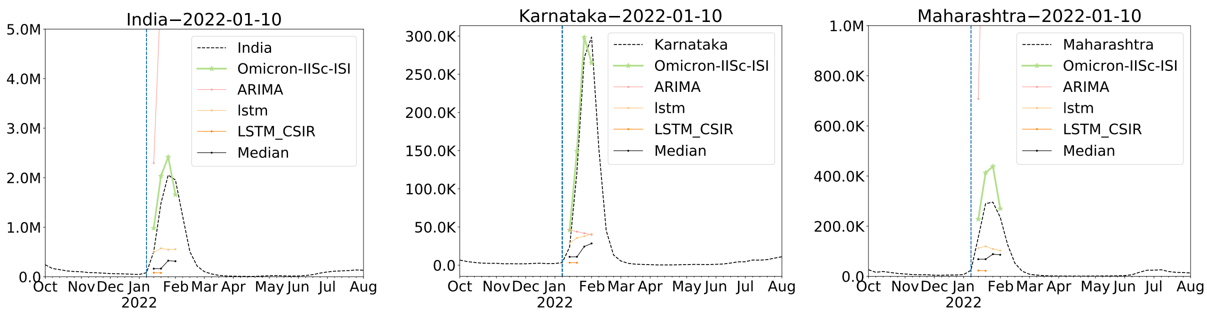
(c) 1-week ahead

(d) 4-week ahead

Figure 2: Performance evaluation of models. (a) Bar plot of mean WIS computed across all locations and forecast weeks for different horizons. (b) Phase classified time series corresponding to cases reported at national level with each week classified into one of three phases. (c) 1-week ahead: Mean WIS computed across all locations for each forecast week. (d) 4-week ahead: Mean WIS computed across all locations for each forecast week. The performance of the Median ensemble is relatively stable and its performance is close to the best performing models across shorter and longer horizons. Also, in order to highlight the performance of models during different phases of the pandemic, we color code weeks in figures (b) and (c) with the corresponding phases. Note that in these results only models forecasting for all states are considered in this evaluation.

In these phases we observe that WIS score, although different across models, are relatively stable. More importantly, we observe that the ensemble has the lowest WIS. January 2022 saw the onset of the *Omicron wave* when cases started to rise sharply. At the onset, the models had the highest WIS score which can be attributed to models' inability to accurately estimate the rapid increase in case counts. But in the subsequent weeks, the models' performances started to improve. Here we observe that the ensemble model, although having relatively poor performance as compared to the previous phases, is able to filter outlier forecasts (note that median is computed per quantile) characterized by high WIS scores. We again observe that during the change of phase from surge to decline, the models' performances declined. After the *Omicron wave*, we saw a drop in the number of submissions from the teams. Due to relatively low numbers of available forecasts, the ensemble model was susceptible to performance variations within individual models.

Forecasting the Omicron wave In order to highlight the importance of simulation models, we consider the example of forecasting the omicron wave. In Figure 3 we show forecasts for three states provided as of 2022-01-10 (a week prior to the onset of the Omicron wave). We observe that the Omicron-IISc-ISI model is able to predict the timing of the onset, peak and decline phase when compared to the other models available at that time. As discussed in Section 3.1.1, the Omicron-IISc-ISI model specifically considers the increased transmissibility of the variant (learnt from South Africa cases data). Once cases related to the variant were detected in a state, the transmissibility (see (1)) is increased from that date onwards and the simulations are run to obtain the forecasts. Since the characteristics of the Omicron wave were different from the other waves, there was no historical data available for the other purely data-driven models to learn the patterns. The simulation model on the other hand allows for the explicit incorporation of the disease parameters thus providing better estimates. Although the simulation model predictions are



(a) India (b) Karnataka (c) Maharashtra

Figure 3: Examples of forecasts of the outbreak due to the Omicron variant for different targets: (a) India, (b) Karnataka, and (c) Maharashtra. The simulation model, Omicron-IISc-ISI captured the onset, peak, and decline parts of the Omicron wave more accurately as compared to the other models.

highly sensitive to the parameters and can lead to over estimation of cases (cf. Figure 3c), the information on the onset of a wave is very helpful for decision makers.

5 RELATED WORKS AND CONCLUSION

Several models emerged from India during the course of the pandemic in order to understand the spread of COVID-19. Simulation models have been particularly useful in capturing the characterstics of the novel virus and provide relatively accurate short-term and long-term predictions. ODE-based models were developed to analyze the effect of travel restrictions (Mandal et al. 2020), mixing across age groups (Singh and Adhikari 2020), social behavioral patterns (Venkateswaran and Damani 2020), and asymptomatic transmissions (Ansumali et al. 2020). A more fine-grained model by Hazra et al. (2022) was used for modeling the first wave by incorporating age stratification, non-pharmaceutical interventions, testing characteristics, etc. Simultaneously, agent-based models were developed in order to understand the spread dynamics at finer resolutions such as city-level under influence of different lockdown relaxation policies (Agrawal et al. 2020), testing strategies (Gopalan and Tyagi 2020), and location-specific lockdowns (Bhattacharyya and Vinay 2020). With the progression of the pandemic, more sophisticated models emerged. Ganesan and Subramani (2021a) developed a high-dimensional partial differential equation-based model to capture the heterogeneous spatio-temporal disease spread dynamics and an ensemble-based variant of the same model was developed later (Ganesan et al. 2021). Foy et al. (2021) developed an age-structured compartmental model with social contact patterns to study the vaccine strategies. Ray et al. (2020) describes a model that resulted in a notable effort towards generating daily forecasts, transmission parameter estimates. These estimates were available for public access at https://covind19.org/.

Among statistical frameworks, exponential models (Ranjan 2020; Gupta and Shankar 2020; Mangla et al. 2021; Singh et al. 2020; Pandey et al. 2022; Ranjan et al. 2021), Poisson regression models (Das 2020), and AR models (Deb and Majumdar 2020; Mangla et al. 2021) were used for case forecasts. Machine learning and deep learning models were experimented with and used for both short- and long-term forecasts (Dukkipati et al. 2021; Bhimala et al. 2022).

Owing to several challenges posed by data quality, policies, and behavior, models tended to lack consensus and importantly failed to predict the Delta wave. Even though there has been extensive development of multiple classes of models to understand the COVID-19 disease spread dynamics in India, a collaborative effort was missing or not well documented.

Ensemble models have shown superior performance as opposed to a single constituent model in several disease forecasting instances, with influenza (Reich et al. 2019), Ebola (Viboud et al. 2018), and dengue fever outbreaks (Johansson et al. 2019) as examples. In the US and Europe, the COVID-19 outbreak prompted the CDC to establish forecast hubs which have seen continued participation from dozens of teams. The results of the analysis from these efforts indicate that ensemble models frequently outperformed individual models (Cramer et al. 2022; Sherratt et al. 2023). The forecasts collated by the hub is documented extensively (Cramer et al. 2022) and is available for public access. In addition, the moderators of the hub also provide interactive visualizations (https://viz.covid19forecasthub.org/, https://covid19forecasthub.eu/visualisation.html)

This paper reports the first attempt in India to develop a *hub* that provided an operational pipeline for real-time submission, aggregation, and communication of forecasts from multiple teams using data-driven AI tools. The need for this framework was motivated by the interaction of our team members with the Karnataka state's COVID-19 TAC. Although the participation in the *hub* was limited, the set of models deployed were relatively diverse. During the course of the effort, new simulation models were developed to address the need of the hour. Enforcing standard submission guidelines enabled the real-time generation of median-based ensemble forecasts and updating of an interactive visualization engine for quick inspection of forecasts. Despite the limited set of models, evaluation results indicate a robust performance of the ensemble through different phases of the pandemic. Throughout the period, the ensemble's performance was at least comparable to the best performing individual models for each particular week.

Although the pipeline was developed for COVID-19 forecasting, it is designed in a pathogen-agnostic manner and can be deployed for forecasting other diseases of concern in India (influenza, dengue, etc.). In terms of ensemble model enhancements, trained ensembles are the immediate choice, where the importance of constituent model's forecast is determined by historical performance. However, the utility of trained ensembles are subject to availability of stable set of models in the training period (Ray et al. 2023).

ACKNOWLEDGMENTS

This work was partially supported by National Science Foundation (NSF) RAPID Grant No. CCF-2142997, University of Virginia Strategic Investment Fund award number SIF160, National Institutes of Health (NIH) Grant 1R01GM109718, NSF Grant No.: OAC-1916805 (CINES), NSF Expeditions in Computing Grant CCF-1918656, NSF RAPID CNS-2028004, NSF RAPID OAC-2027541, Gift from Google, LLC, Indian Institute of Science IoE grant, Centre for Networked Intelligence grant, and ICTS Knowledge Exchange grant, Gift from the William H. Miller III 2018 Trust.

REFERENCES

Adiga, A., D. Dubhashi, B. Lewis, M. Marathe, S. Venkatramanan, and A. Vullikanti. 2020. "Mathematical Models for COVID-19 Pandemic: A Comparative Analysis". *Journal of the Indian Institute of Science* 100(4):793–807.

Adiga, A., G. Kaur, B. Hurt, L. Wang, P. Porebski, S. Venkatramanan, B. Lewis, and M. Marathe. 2022. "Enhancing COVID-19 Ensemble Forecasting Model Performance Using Auxiliary Data Sources". In *2022 IEEE International Conference on Big Data (Big Data)*, 1594–1603. IEEE.

Agrawal, M., M. Kanitkar, and M. Vidyasagar. 2021. "SUTRA: An Approach to Modelling Pandemics with Undetected (Asymptomatic) Patients, and Applications to COVID-19". In *2021 60th IEEE Conference on Decision and Control (CDC)*, 3531–3531.

Agrawal, S., S. Bhandari, A. Bhattacharjee, A. Deo, N. M. Dixit, P. Harsha, S. Juneja, P. Kesarwani, A. K. Swamy, P. Patil et al. 2020. "City-Scale Agent-Based Simulators for the Study of Non-Pharmaceutical Interventions in the Context of the COVID-19 Epidemic: IISc-TIFR COVID-19 City-Scale Simulation Team". *Journal of the Indian Institute of Science* 100:809–847.

Ansumali, S., S. Kaushal, A. Kumar, M. K. Prakash, and M. Vidyasagar. 2020. "Modelling a Pandemic With Asymptomatic Patients, Impact of Lockdown and Herd Immunity, with Applications to SARS-CoV-2". *Annual Reviews in Control* 50:432–447.

Bezjak, S., A. Clyburne-Sherin, P. Conzett, P. L. Fernandes, E. Görögh, K. Helbig, B. Kramer, I. Labastida, K. Niemeyer, F. Psomopoulos et al. 2018. "The Open Science Training Handbook". Technical report, [sn].

Bhattacharyya, C., and V. Vinay. 2020. "Suppress, and Not Just Flatten: Strategies for Rapid Suppression of COVID19 Transmission in Small World Communities". *Journal of the Indian Institute of Science* 100:849–862.

Bhimala, K. R., G. K. Patra, R. Mopuri, and S. R. Mutheneni. 2022. "Prediction of COVID-19 Cases Using the Weather Integrated Deep Learning Approach for India". *Transboundary and Emerging Diseases* 69(3):1349–1363.

Bracher, J., E. L. Ray, T. Gneiting, and N. G. Reich. 2021. "Evaluating Epidemic Forecasts in an Interval Format". *PLoS Computational Biology* 17(2):e1008618.

Cramer, E. Y., Y. Huang, Y. Wang, E. L. Ray, M. Cornell, J. Bracher, A. Brennen, A. J. C. Rivadeneira, A. Gerding, K. House et al. 2022. "The United States COVID-19 Forecast Hub Dataset". *Scientific Data* 9(1):462.

Cramer, E. Y., E. L. Ray, V. K. Lopez, J. Bracher, A. Brennen, A. J. Castro Rivadeneira, A. Gerding, T. Gneiting, K. H. House, Y. Huang et al. 2022. "Evaluation of Individual and Ensemble Probabilistic Forecasts of COVID-19 Mortality in the United States". *Proceedings of the National Academy of Sciences* 119(15):e2113561119.

Das, S. 2020. "Prediction of COVID-19 Disease Progression in India: Under the Effect of National Lockdown". *arXiv preprint arXiv:2004.03147*.

Deb, S., and M. Majumdar. 2020. "A Time Series Method To Analyze Incidence Pattern and Estimate Reproduction Number of COVID-19". *arXiv preprint arXiv:2003.10655*.

Doms, C., S. C. Kramer, and J. Shaman. 2018. "Assessing the Use of Influenza Forecasts and Epidemiological Modeling in Public Health Decision Making in the United States". *Scientific Reports* 8(1):12406.

Dukkipati, A., T. Gracious, and S. Gupta. 2021. "CoviHawkes: Temporal Point Process and Deep Learning Based Covid-19 Forecasting for India". *arXiv preprint arXiv:2109.06056*.

Foy, B. H., B. Wahl, K. Mehta, A. Shet, G. I. Menon, and C. Britto. 2021. "Comparing COVID-19 Vaccine Allocation Strategies in India: A Mathematical Modelling Study". *International Journal of Infectious Diseases* 103:431–438.

Ganesan, S., and D. Subramani. 2021a. "Spatio-Temporal Predictive Modeling Framework for Infectious Disease Spread". *Scientific Reports* 11(1):1–8.

Ganesan, S., and D. Subramani. 2021b. "Why We Need Ensemble COVID Predictions". *The New Indian Express*. https://www.newindianexpress.com/opinions/2021/jun/01/why-we-need-ensemble-covid-predictions-2310039.html.

Ganesan, S., D. Subramani, T. Anandh, D. Ghose, and G. R. Babu. 2021. "Ensemble Forecast of COVID-19 in Karnataka for Vulnerability Assessment and Policy Interventions". *medRxiv*:2021–08.

Gopalan, A., and H. Tyagi. 2020. "How Reliable Are Test Numbers for Revealing the COVID-19 Ground Truth and Applying Interventions?". *Journal of the Indian Institute of Science* 100:863–884.

Gupta, S., and R. Shankar. 2020. "Estimating the Number of COVID-19 Infections in Indian Hot-Spots Using Fatality Data". *arXiv preprint arXiv:2004.04025*.

Hazra, D. K., B. S. Pujari, S. M. Shekatkar, F. Mozaffer, S. Sinha, V. Guttal, P. Chaudhuri, and G. I. Menon. 2022. "Modelling the First Wave of COVID-19 in India". *PLoS Computational Biology* 18(10):e1010632.

Hochreiter, S., and J. Schmidhuber. 1997. "Long Short-Term Memory". *Neural Computation* 9(8):1735–1780.

Howerton, E., M. C. Runge, T. L. Bogich, R. K. Borchering, H. Inamine, J. Lessler, L. C. Mullany, W. J. Probert, C. P. Smith, S. Truelove et al. 2023. "Context-Dependent Representation of Within-and Between-Model Uncertainty: Aggregating Probabilistic Predictions in Infectious Disease Epidemiology". *Journal of the Royal Society Interface* 20(198):20220659.

Johansson, M. A., K. M. Apfeldorf, S. Dobson, J. Devita, A. L. Buczak, B. Baugher, L. J. Moniz, T. Bagley, S. M. Babin, E. Guven et al. 2019. "An Open Challenge to Advance Probabilistic Forecasting for Dengue Epidemics". *Proceedings of the National Academy of Sciences* 116(48):24268–24274.

Lutz, C. S., M. P. Huynh, M. Schroeder, S. Anyatonwu, F. S. Dahlgren, G. Danyluk, D. Fernandez, S. K. Greene, N. Kipshidze, L. Liu et al. 2019. "Applying Infectious Disease Forecasting to Public Health: A Path Forward Using Influenza Forecasting Examples". *BMC Public Health* 19(1):1–12.

Mandal, S., T. Bhatnagar, N. Arinaminpathy, A. Agarwal, A. Chowdhury, M. Murhekar, R. R. Gangakhedkar, and S. Sarkar. 2020. "Prudent Public Health Intervention Strategies to Control the Coronavirus Disease 2019 Transmission in India: A Mathematical Model-Based Approach". *The Indian Journal of Medical Research* 151(2-3):190.

Mangla, S., A. K. Pathak, M. Arshad, and U. Haque. 2021. "Short-Term Forecasting of the COVID-19 Outbreak in India". *International Health* 13(5):410–420.

Pandey, P., J. Gómez-Aguilar, M. K. Kaabar, Z. Siri, and A. M. Abd Allah. 2022. "Mathematical Modeling of COVID-19 Pandemic in India Using Caputo-Fabrizio Fractional Derivative". *Computers in Biology and Medicine* 145:105518.

Ranjan, R. 2020. "Predictions for COVID-19 Outbreak in India Using Epidemiological Models". *medRxiv*:2020–04.

Ranjan, R., A. Sharma, and M. K. Verma. 2021. "Characterization of the Second Wave of COVID-19 in India". *medRxiv*:2021–04.

Ray, D., M. Salvatore, R. Bhattacharyya, L. Wang, J. Du, S. Mohammed, S. Purkayastha, A. Halder, A. Rix, D. Barker et al. 2020. "Predictions, Role of Interventions and Effects of a Historic National Lockdown in India's Response to the COVID-19 Pandemic: Data Science Call to Arms". *Harvard Data Science Review* 2020(Suppl 1).

Ray, E. et al. 2021. "Challenges in Training Ensembles To Forecast COVID-19 Cases and Deaths in the United States". *International Institute of Forecasters*.

Ray, E. L., L. C. Brooks, J. Bien, M. Biggerstaff, N. I. Bosse, J. Bracher, E. Y. Cramer, S. Funk, A. Gerding, M. A. Johansson et al. 2023. "Comparing Trained and Untrained Probabilistic Ensemble Forecasts of COVID-19 Cases and Deaths in the United States". *International Journal of Forecasting* 39(3):1366–1383.

Reich, N. G., L. C. Brooks, S. J. Fox, S. Kandula, C. J. McGowan, E. Moore, D. Osthus, E. L. Ray, A. Tushar, T. K. Yamana et al. 2019. “A Collaborative Multiyear, Multimodel Assessment of Seasonal Influenza Forecasting in the United States”. *Proceedings of the National Academy of Sciences* 116(8):3146–3154.

Sherratt, K., H. Gruson, H. Johnson, R. Niehus, B. Prasse, F. Sandmann, J. Deuschel, D. Wolffram, S. Abbott, A. Ullrich et al. 2023. “Predictive Performance of Multi-Model Ensemble Forecasts of COVID-19 Across European Nations”. *Elife* 12:e81916.

Singh, R., and R. Adhikari. 2020. “Age-Structured Impact of Social Distancing on the COVID-19 Epidemic in India”. *arXiv preprint arXiv:2003.12055*.

Singh, R. K., M. Drews, M. De La Sen, M. Kumar, S. S. Singh, A. K. Pandey, P. K. Srivastava, M. Dobriyal, M. Rani, P. Kumari et al. 2020. “Short-Term Statistical Forecasts of COVID-19 Infections in India”. *IEEE Access* 8:186932–186938.

Venkateswaran, J., and O. Damani. 2020. “Effectiveness of Testing, Tracing, Social Distancing and Hygiene in Tackling COVID-19 in India: A System Dynamics Model”. *arXiv preprint arXiv:2004.08859*.

Viboud, C., K. Sun, R. Gaffey, M. Ajelli, L. Fumanelli, S. Merler, Q. Zhang, G. Chowell, L. Simonsen, A. Vespignani et al. 2018. “The RAPIDD Ebola Forecasting Challenge: Synthesis and Lessons Learnt”. *Epidemics* 22:13–21.

Yamana, T. K., S. Kandula, and J. Shaman. 2016. “Superensemble Forecasts of Dengue Outbreaks”. *Journal of The Royal Society Interface* 13(123):20160410.

AUTHOR BIOGRAPHIES

ANIRUDDHA ADIGA is a Research Scientist at Biocomplexity Institute, aniruddha@virginia.edu.

SIVA ATHREYA is a professor at the Indian Statistical Institute, Bangalore Centre and the International Centre for Theoretical Sciences-TIFR, athreya@icts.res.in.

KANTHA RAO BHIMALA is a principal scientist at CSIR Fourth Paradigm Institute, Bengaluru, kantha@csir4pi.in.

AMBEDKAR DUKKIPATI is a professor at Computer Science and Automation, Indian Institute of Science, ambedkar@iisc.ac.in.

TONY GRACIOUS is a PhD student at the Indian Institute of Science, tonygracious@iisc.ac.in.

SHUBHAM GUPTA was a PhD student at the Indian Institute of Science. He is now with IBM Paris, shubhamg@iisc.ac.in.

BENJAMIN HURT is a Data Scientist at the Biocomplexity Institute, benjaminhurt@virginia.edu.

GURSHARN KAUR is a postdoctoral research associate at the Biocomplexity Institute, gursharn@virginia.edu.

BRYAN LEWIS is a research associate professor at the Biocomplexity Institute, bl4zc@virginia.edu.

MADHAV MARATHE is a Distinguished Professor at the Biocomplexity Institute, marathe@virginia.edu.

VIDYADHAR MUDKAVI is a distinguished Emeritus Professor at the CSIR Fourth Paradigm Institute, Bengaluru, vidyadhar.mudkavi@gmail.com.

GOPAL KRISHNA PATRA is a Chief Scientist at the CSIR Fourth Paradigm Institute, Bengaluru, gkpatra@csir4pi.in.

PRZEMYSLAW POREBSKI is a Research Scientist at Biocomplexity Institute, pjp2b@virginia.edu.

NIHESH RATHOD was a PhD student at the Indian Institute of Science. He is currently with Strand Life Sciences, nihesh.7391@gmail.com.

RAJESH SUNDARESAN is a professor at Electrical Communication Engineering, Indian Institute of Science, rajeshs@iisc.ac.in.

SRINIVASAN VENKATRAMANAN is a research assistant professor at the Biocomplexity Institute, sv8nv@virginia.edu.

SARATH YASODHARAN was a PhD student at IISc when this work was done. He is currently a post-doctoral fellow at Brown University, sarath1789@gmail.com.