Forecasting the spread of COVID-19

Sharad Chitlangia[§]
Department of Electrical and Electronics Engineering
BITS Pilani, K.K. Birla Goa Campus
Goa 403726, India
f20170472@goa.bits-pilani.ac.in

Devanshu Shah[§]
Department of Computer Science and Information Systems
BITS Pilani, K.K. Birla Goa Campus
Goa 403726, India
f20180240@goa.bits-pilani.ac.in

Vybhav Jayasankar[§]
Department of Computer Science and Information Systems
BITS Pilani, K.K. Birla Goa Campus
Goa 403726, India
f20180152@goa.bits-pilani.ac.in

Asad Hazariwala[§]
Department of Computer Science and Information Systems
BITS Pilani, K.K. Birla Goa Campus
Goa 403726, India
f20170774@goa.bits-pilani.ac.in

***Abstract*—We tackle the problem of time series forecasting of COVID-19 daily case counts through Machine Learning Methods. By modelling the COVID-19 daily case counts with autoregressive processes, we show that even simple regression based methods fare as well as traditional compartmental epidemic models. To utilize the expressiveness of Deep Learning Frameworks, we develop a framework around Meta Learning in a Few Shot Time Series Forecasting setting. With our framework, we show that using only a few datapoints, and in only a single gradient step, a meta learning based model can predict as well as, or at times, even better than a model trained on country specific data for hundreds of epochs. We thoroughly analyse the choice of our features in this study by quantifying the Granger Causal relationship between features and daily case counts. Deeper analysis uncovers interesting insights about Government Interventional Policies, and the effects of recent events such as Kumbh Mela in Uttarakhand and the election rallies held in Tamil Nadu and West Bengal.**

***Index Terms*—COVID-19, Machine Learning, Causal Inference, Time Series Forecasting**

I. Introduction

Since the beginning of 2020, the coronavirus disease (COVID-19) has had a devastating impact on human society. From disrupting travel, the downfall of whole economies to stricter social rules, the human race has fought hard to prevent its spread. Even with massive vaccination campaigns and stringent regulations, the coronavirus is nowhere to go as yet. To our dismay and perhaps due to slow responses, the coronavirus has also evolved into multiple different strains from the variant originating in Wuhan, China. Different variants of the virus have been now been detected in UK and India against which the original vaccines are not as effective [1]. As an example of the surge of cases due to this, India is now facing its 2nd wave with more than 400K cases and 3k deaths everyday at the time of writing.

It is important that with such implications, the spread of the disease be forecasted. This is due to multiple reasons: 1) Forecasting the spread of the disease helps health authorities to keep in check the health resources available for use (In particular, plan for the future efficiently). 2) Aid the government in making decisions on administrative interventional policies such as lockdown, travel restrictions, etc., aimed at containing the spread of the disease.

Even though India boasts one of the highest vaccine manufacturing capabilities (through the Serum Institute, Pune) and also one of the cheapest health services across the world, India's health system is now on the verge of collapse due to the sheer load of the total number of active cases, which is estimated to be around 2600K. The new surge is attributed, by experts [1], to the slow and poor response in part by the government. Recent religious events such as the Kumbh Mela, election rallies in several states leading up to state elections and in general, complacency among people have also been portrayed as the major reasons for this new surge of cases.

In such times requiring strong forecasting models, epidemiologists often rely on 'compartmental models' for epidemic forecasting, in particular, SIR and SEIR. These models keep track of the status of infections, recoveries and deaths to compartmentalize the population. The dynamics of spread are then modelled using differential equations using these variables. Although these models provide a rough estimate, the variables considered are often uncontrollable and influenced by a variety of real world factors. This often makes these models somewhat unreliable, which is of utmost importance. An example of this is the failure of the SUTRA [2] model by a team of scientists chosen by the Department of Science and Technology, India. It builds on SIR to account for asymptomatic patients and has recently failed to predict the second wave which seemingly has not reached its peak. Additionally, such models also do not account for and encode real world covariates such as mobility, twitter activity, etc., which could be really important signals in forecasting the spread of COVID-19.

This brings us to the question of whether Machine Learning

[§]Equal contribution

[1]https://thewire.in/health/five-covid-19-policy-mistakes-india-could-have-done-without

models can outperform such methods? The primary advantage to choosing data driven Machine Learning models are that these models are often very *robust* to real world data. Case Counts, Death Counts, Positivity Ratio are all crowdsourced datapoints and hence naturally contain some amount of noise. But even though there are benefits to choosing standard Machine Learning based time series models (such as Sequence to Sequence models), they often require millions of datapoints to forecast robustly. For comparison, the length of the time series of COVID-19 daily counts is only 450. This raises the important question of whether it is possible to devise deep learning based methods that can robustly forecast the daily COVID-19 case counts using only a few datapoints from that time series?

The spread of COVID-19 or the disease dynamics, is very dependent on the movement of individuals. For instance, some epidemiologists hypothesize that Italy was one of the first to be hit with COVID-19 after China because of the tourist movement from China to Italy [3] [4]. However, we note that most governments impose interventional policies such as lockdowns, work from home policies, etc., to limit movement. This means that only essential services are active after these policies are enforced. Interventional policies and their regulations over the past year and a half, can be analysed on the basis of how stringent the policies governments have introduced. After such a policy is imposed, it is a fairly good assumption to treat the dynamics of that region independent of any other. What this means is that, for a large number of demarcated regions, the relative evolution of the disease might remain the same. Motivated by this observation, we hypothesize that the problem of univariate forecasting (with or without exogenous variables such as interventional policies) can be treated in a 'Meta-Learning' framework to transfer learn across countries or regions. We propose that each demarcated region has its own dynamics, but in the larger scheme of things, there are a lot of temporally correlated factors that remain common across regions that can be learnt. After Meta-Training and fine tuning on only a few time series datapoints, predictions about the case counts in that demarcated region can be made very accurately. This means that using our proposed framework, any demarcated region may be able to forecast its daily counts very robustly and accurately after utilizing the knowledge of disease dynamics of other regions.

It is very important to choose the right features as inputs to our models and hence, to evaluate our choices, we study how particular features could be helpful in forecasting from a causal perspective. For this we propose to use the Granger causality definitions as proposed in [5]. In the context of time series, what a causal feature means is how well it helps in forecasting in the future. That is, how 'temporally correlated' two particular sequences may be. We limit our analysis to linear models to avoid latent factors from playing any role. This is because it becomes harder to quantify latent cause effect relationships especially in the case of time series models. Through our analysis we show which mobility metrics are actually useful and which are not. We also show how the recent spread of the COVID-19 virus in some districts can be attributed to higher case counts in cities where recent events such as election rallies or Kumbh Mela took place.

Our contributions are as follows:

1) We demonstrate the applicability of a range of Machine Learning Models in forecasting the spread of Coronavirus. These models perform either better than or as well as Compartmental Epidemic Forecasting models.
2) We present a novel framework for few shot time series forecasting of daily COVID-19 cases which transfer learns on mobility and stringency patterns using Meta Learning to enable the usage of Deep Learning models.
3) We quantify and demonstrate the existence of causal relationships between particular variables such as mobility, stringency and daily case counts.
4) We study in detail recent events such as the Kumbh Mela and the election rallies in Tamil Nadu. In particular, through causal analysis, we verify the spread of coronavirus through certain hotspot regions where these events took place in the case of the Kumbh Mela.

The remainder of the paper is structured as follows: Section II provides a brief overview of the related work, Section III starts by describing the formulation and introduces the different methods we use in this paper. Section IV goes over some of the details about the data, the core questions being investigated, and the materials utilized. In Section V, we present our results and some analysis. In Section VI, we conclude by summarizing and describing some of the future work possible.

II. Related Work

Since the onset of the coronavirus pandemic, the research community has studied the spread of the virus in different locales, and attempted to forecast the same. Among other things, researchers have used the underlying dynamics of the disease, mobility patterns of people and policy measures to understand the spread of the virus and subsequently forecast the same.

A. Compartmental Models for COVID Forecasting

Chang et al. [6] use a SEIR (Susceptible, Exposed, Infectious and Recovered) based model which integrates dynamic mobility networks, to simulate the spread of COVID-19 in the US. They identify select hotspots for the spread of the virus, and advocate for policy responses aimed at these specific locations, rather than uniform restrictions.

Bedi et al. [7] use a SEIRD (Susceptible, Exposed, Infectious, Recovered and Dead) model to forecast the long-term spread of COVID-19 in India, and an LSTM based model to make short-term predictions. They also analyse the efficacy of the various phases of lockdown enforced by the Indian government by studying the variation in the parameters of the SEIRD model during each of these phases.

Agarwal et. al. [2] proposed a model for the Indian subcontinent that also accounts for asymptomatic cases. But the proposed model has failed to predict the second wave currently ravaging India.

B. Machine Learning Models for COVID Forecasting

Previously, Supervised Learning models have been considered for the problem of time series forecasting. Tomar and Gupta [8] use an LSTM (Long Short Term Memory) [9] model to forecast the number of cases of COVID-19 in India over a 30-day period. The authors have also analysed the efficacy of government lockdown measures, using the transmission rate of the virus as a metric, and found that the early preventive measures were fairly successful in controlling the spread of the virus.

C. Intra Regional Modelling

Kapoor et al. [10] describe a GNN [11] based approach to forecast the incidence of COVID-19 in America, on a county scale. Their model utilises both inter-region and intra-region mobility information, without making any assumptions about the underlying disease dynamics.

Le et al. [2] proposed a Neural Relational Autoregression based model which accounts for interregional features through vector autoregression. They utilize LSTM [9] to provide either a dampening or strengthening factor to the case counts which are finally modelled using a Negative Binomial Distribution.

D. Usage of Mobility Data

Praharaj and Han [12] analysed the effect of various categories of mobility (from the Google Community Mobility Reports) on the incidence of COVID-19 in India. They used a generalized estimating equation with a Poisson log-linear model, and found that certain types of mobility are more significantly associated with the spread of the virus. They proposed that restrictions aimed at these specific areas might be more effective than a blanket lockdown.

Facebook AI Research has also released Movement Range Maps [3] under their Data for Good initiative [4] which are particularly more information than Google Mobility Trends. Unfortunately, the movement range maps does not have any data for the Indian subcontinent where we keep our focus in this paper.

E. Transfer Learning for COVID-19

Many researchers have delved into utilizing transfer learning, especially in the case of predicting COVID-19 on Chest X Ray Images [13]. But research on transfer learning for forecasting is still very limited. [14] transfer learn a standard LSTM [9] based model on cases and deaths.

Li et al. [15] extended this to utilizing stringency measures. They utilize an Attention based RNN approach to encode these features to show that transfer learning is indeed viable. In contrast to this, we utilize a state of the art model 'N-beats' and propose to use Meta Learning to fine tune with a single gradient step of a particular country's data. We also utilize mobility data as opposed to only stringency measures.

F. Causal Analysis of Mobility and Stringency Measures

Although causal analysis of Mobility and stringency measures exists [16], [17], none of them to best of our knowledge exist in the case of the Indian sub-continent. Additionally, we also analyse and attribute some of the recent spikes seen to these events by showcasing how certain regions acted as hotspots for COVID-19 spread.

III. Methodology

A. Problem Formulation

We assume forecasting the spread of the coronavirus disease as a time series forecasting problem by modelling 'daily new cases'. In particular, given $\mathcal{Y} = (y_1, \ldots, y_t)$ of t time steps, our aim is to forecast $(y_{t+1}, \ldots, y_{t+f})$, where f denotes the number of forecast days. At any time, the disease dynamics are constantly affected by various external factors, which we consider as 'exogenous' variables. These variables include interventional government policies, human mobility information, etc., which are also time varying processes.

B. Baselines

1) Persistence Algorithm: The persistence algorithm assumes the forecast of the next f days to be the last observed day's value. Although this model may not be realistic, it serves as a strong baseline for us to evaluate our approach.

$$y_{t+i} = y_t, \text{where } i \in \{1, \ldots, f\} \tag{1}$$

2) Generalized Linear Models (GLM): As the name suggests, GLMs are a generalization of the traditional regression models which are able to model the response variable with an arbitrary distribution. To allow for regression models to work, we assume that daily case counts can be treated as an autoregressive process with an order p. This means that if C_t denotes the case counts on a particular day, then:

$$C_t = \sum_{k=1}^{p+1} (a_k C_{t-k}) \tag{2}$$

Since daily COVID cases is count data, a natural assumption is to model the response variable to be from a Poisson Distribution.

$$Y_t | X_t \sim Poisson(\lambda_t) \tag{3}$$

Here X_t is the combined set of exogenous features and previous day counts. But daily COVID-19 cases exhibits high dispersion from the mean and hence modelling the response variable from the Poisson Distribution may not be expressible enough. Instead, we utilize the Negative Binomial Distribution which allows count data to be modelled while also allowing the variance to be different from the mean.

$$Y_t | X_t \sim NB(\lambda_t, \lambda_t + \alpha \lambda_t^2) \tag{4}$$

[2]https://ai.facebook.com/research/publications/neural-relational-autoregression-for-high-resolution-covid-19-forecasting

[3]https://dataforgood.fb.com/tools/movement-range-maps/

[4]https://dataforgood.fb.com/

Both these models are learnt by minimizing the negative likelihood. We also utilize Elastic Net Regularization, which is a combination of L1 and L2 regularization i.e., if the overall **penalty** term is $L_{penalty}$ and the model parameters can be denoted with β then,

$$L_{penalty} = \lambda_1\|\beta\|_1 \ + \ \lambda_2\|\beta\|_2 \tag{5}$$

and the overall objective is

$$L \ = \ -log(likelihood) \ + \ L_{penalty} \tag{6}$$

3) Long Short Term Memory (LSTM): LSTMs [9] are a type of artificial recurrent neural network which are particularly effective in processing time series data due to their ability to ascertain order in sequence prediction problems. The internal structure of a LSTM was designed specifically to overcome the vanishing gradient problem faced by traditional RNNs, and usually consists of a cell, an input gate, an output gate and a forget gate.

We utilise a model which has a LSTM layer followed by a single dense layer, which given a sequence of case counts, returns the predicted case count for the next day. This prediction can now be fed back into the model, enabling the model to make auto-regressive forecasts of the desired length. Consider the following time series, $\mathbf{Y} = \{y_1, y_2, \ldots, y_n\}$ where $y_i \in R^d$ We split the above series into P disjoint training chunks and Q disjoint testing chunks. Each of those P chunks are of the $2D$ tensors of size $t+1 \times d$, where we see t days and predict the $t+1^{th}$ day. We stack these to create a 3D tensor of the shape $(P \times (t+1) \times d)$. Each of the Q chunks are $2D$ tensors of size $t+f \times d$, where we see t days and predict the next f days. We stack these to create a 3D tensor of the shape $(Q \times (t+f) \times d)$.

During training, the model predicts only the next day, i.e. $t+1^{th}$ day, because of the nature of the LSTM. However during the testing phase, the model predicts for the next f days in an autoregressive manner, i.e. it predicts for the $t+1^{th}$ day and then we remove the 1^{st} day and add the $t+1^{th}$ for the length of the input to still be t. This is done f times to get the prediction for the next f days.

C. Compartmental Model

We consider a discrete SIR(Susceptible-Infected-Recovered) model. We assume the disease spreads at a rate of λ from infected people (I) to susceptible people (S) and the infected people recover at a rate of μ to become recovered people (R).

$$\begin{aligned} S + I &\xrightarrow{\lambda} I + I \\ I &\xrightarrow{\mu} R \end{aligned} \tag{7}$$

The following equations model the above situation.

$$\begin{aligned} \frac{dS}{dt} &= -\lambda\frac{SI}{N} \\ \frac{dI}{dt} &= \lambda\frac{SI}{N} + \mu I \\ \frac{dR}{dt} &= -\mu I \end{aligned} \tag{8}$$

Since we have a discrete dataset, we approximate (8) as follows:

$$\begin{aligned} S_t - S_{t-1} &= -\lambda\Delta t\frac{S_{t-1}I_{t-1}}{N} \\ &=: -I_t^{new} \\ R_t - R_{t-1} &= -\mu\Delta t I_{t-1} \\ &=: R_t^{new} \\ I_t - I_{t-1} &= (\lambda\frac{S_{t-1}}{N} - \mu)\Delta t I_{t-1} \\ &=: I_t^{new} - R_t^{new} \end{aligned} \tag{9}$$

An important point to note is that I_t models the number of active (currently) infected people, while I_t^{new} is the number of daily new infections that are reported according to the MoHFW, India. Moreover we also add an explicit term to account for the delay D between new infections and reported cases when generating the forecast.

1) Estimating model parameters: We estimate the model parameters $\theta = \{\lambda, \mu, \sigma, I_0\}$ using Bayesian inference.

1) Choose random initial parameters based on an explicitly specified prior distribution. Then time integration of the model generates a fully deterministic time series of the new cases $I^{new}(\theta)$ of the same length as the observed data.
2) Recursively update parameters and the time integration in every MCMC step. We quantify the difference in the model outcome $I_t^{new}(\theta)$ and the real data $\hat{I}_t^{new}$ with a *StudentT*

$$\hat{I}_t^{new}|\theta \sim \\ StudentT_{\nu=4}(mean = I_t^{new}(\theta), scale = \sigma\sqrt{I_t^{new}(\theta)})) \tag{10}$$

3) For forecasting we take all the MCMC samples and again run the time integration on them, to forecast the new cases. Note that we don't just use the maximum A Posteriori estimates of parameters for the forecasting but rather the MCMC samples and hence we can model the uncertainty in the forecasts.

Specifying Priors

$$\begin{aligned} \lambda &\sim LogNormal(log(0.4), 0.5) \\ \mu &\sim LogNormal(log(1/8), 0.2) \\ \sigma &\sim HalfCauchy(1) \\ I_0 &\sim LogNormal(log(\hat{I}_0), 0.9) \end{aligned} \tag{11}$$

D. ARIMA

Auto Regressive Integrated Moving Average (ARIMA) is a model commonly used in time-series analysis which incorporates features of auto regressive (AR) and moving average models (MA). In its most general form, an ARIMA model is specified by 3 non-negative parameters (p,d,q), where p is the order of the auto-regressive model, q is the order of the moving average model and d is the number of times the data is being differenced.

ARIMA models are particularly suited to non-seasonal time-series data in which the mean exhibits non-stationary tendencies, while the variance and autocovariance are stationary. ARIMA deals with the non-stationarity in the mean by differencing the data. The mean of the daily case counts being dispersed and the data not exhibiting seasonal tendencies motivate us to model the case counts using ARIMA. We do not incorporate any exogenous variables in our model, though this is possible within the ARIMA framework.

Given a time series $X = (X_1, X_2, ..)$, we define a backward shift operator L as :

$$\begin{aligned} LX_t &= X_{t-1} \; \forall \; t > 1 \\ L^k X_t &= X_{t-k} \end{aligned} \tag{12}$$

Let C_t denote the number of cases reported on a particular day. An ARIMA model(p,d,q) is summarized by the following equation :

$$(1 - \sum_{i=1}^{p} \alpha_i L^i)(1-L)^d C_t = \delta + (1 + \sum_{j=1}^{q} \beta_j L^j)\epsilon_t \tag{13}$$

The parameters α_i and β_j in (13) are estimated using a maximum likelihood approach and the order of the model is selected so as to minimize Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC). In our case, we found the minimum AIC and BIC for $p = 7$, $d = 1$, $q = 0$. A drawback to using ARIMA to model cases is that the variance of the data is not stationary throughout the entire time period under consideration, which could potentially affect model performance.

E. Prophet

Prophet is an open-source library developed by Facebook [18], for the purpose of fast, automatic univariate time-series forecasting. It is a very powerful and widely used tool for time-series forecasting, and as such, serves as a good baseline for our problem.

Prophet uses a decomposable time series model, whose 3 main components are trend, seasonality and holidays, represented in the following equation :

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{14}$$

where $g(t)$ is the trend function, $s(t)$ represents the seasonality, $h(t)$ represents the effect of holidays and ϵ_t is an error term, which is assumed to be normal. Prophet does not explicitly consider the temporal dependence structure in a time-series problem, but rather treats it as a curve-fitting problem.

Prophet supports two trend models, a non-linear saturating growth model which is modelled using a logistic growth model, and a linear trend model intended for problems without saturating growth. We utilise the linear trend model here, as we have yet to see a saturation in COVID-19 cases, and it is difficult to say what that limit may be. The effect of the seasonality is also added to the trend when making a forecast.

In forecasting the trend, Prophet assumes that the future will see similar trend changes as seen in the past. The trend model for the seen data consists of S changepoints out of T points, each of which has a rate change $\delta_j \sim Laplace(0, \tau)$. To predict future trend changes, τ is replaced with a variance inferred from the seen data.

F. N-beats

N-beats [19] is a recently proposed pure deep learning based architecture which treats the problem of time series forecasting as that of predicting basis coefficients of the time series. This allows it to arbitrarily work with any time series data, as scaling the inputs (e.g., to remove non stationary trends) is not needed. This is one of the reasons we pick N-beats as our model for Meta-Learning, because it allows us to learn across COVID count time series without requiring any country specific normalization.

G. Few Shot Time Series Forecasting via Meta Learning

Univariate Time Series Forecasting on COVID count data with deep learning based models is particularly hard as it only has at most, 500 days worth of datapoints. A way around this is to consider overlapping sequences as training samples. But this may result in the model observing particular trends more often than others. We propose to get around this issue by utilizing recent progress in Meta Learning literature. In particular, we propose that by learning trend patterns across geographical areas as independent tasks through Meta Learning, univariate time series forecasting is made tractable. In a gist, Meta Learning aims to learn good initialization so as to learn new unseen tasks with very less samples. It does so by iteratively learning a common model across task distributions so that it can fine-tuned for a particular task with only a few gradient steps.

We utilize REPTILE, which is an optimization based meta learning algorithm. Training is performed over different demarcated regions and is finally tested by fine tuning on the final region. For instance, for country level forecasts we first perform training over COVID Case Counts of multiple different countries and fine tune with a few time series samples on India to forecast.

Algorithm 1 REPTILE Batched Version

```
Initialize θ
for iteration = 1, 2, ... do
    Sample tasks τ_1, τ_2, ..., τ_n
    for i = 1, 2, ...n do
        Compute W_i = SGD(L_{τ_i}, θ, k)
    end for
    Update θ ⟵ θ + β (1/n) Σ_{i=1}^{n} (W_i, θ)
end for
```

H. Granger Causality

For time series data, a feature may be considered to be 'causal' if it helps in 'forecasting'. That is, for unseen data, if the inclusion of a feature reduces the variance in prediction error then it can be treated as a causal feature. Consider two time series $C_t = \{C_1, C_2, ...C_T\}$ and $D_t = \{D_1, D_2, ..D_T\}$. We assume each time series to be an autoregressive process of order p.

$$C_{t+1} = \sum_{k=1}^{p+1}(a_k C_{t-p}) + \sum_{k=1}^{p+1}(b_k D_{t-p}) + E_1(t) \quad (15)$$

The equation 15 desribes a Vector Autoregression (VAR) model. VAR models study the evolution of multiple time series (as endogenous variables) together. We use VAR to study Granger Causality. If by inclusion of the time series D, the variance of $E_1(t)$ is reduced then we say that D_t *Granger Causes* C_t. Alternatively, this means that if the set of coefficients $\{b_k\}$ of D_t in equation (15) are not 0 then, D_t *Granger Causes* C_t. For more precise definitions of Granger Causality, we refer the reader to its original definitions in [5]. In short, we perform an F-test with the null hypothesis (H_0) that $b_k = 0$. We set the significance level to be 5%. If we reject this hypothesis then we can conclude that D_t indeed *Granger Causes* C_t.

IV. Empirical Evaluation

A. Materials

Data Source	Description
Our World in Data	Aggregate country level data with multiple features (cases, deaths, tests, etc.)
COVID-19 India API	Aggregated State and district level case counts
Google Community Mobility Reports	Country, state and district level time series data quantifying mobility statistics
Apple Mobility Trends Reports	Mobility data for the last year which reflects requests for directions in Apple Maps

TABLE I: A summary of the data sources used in this paper, along with a description of the data obtained from each source.

1) Data: We utilize publicly available data from various sources which are summarized in I. The data at these sources is aggregated based on media reports from government portals[5].

For our experimentation, we primarily utilize case counts. Case counts are available at the district level and above. To encode external factors as exogenous variables, we utilize the time series of stringency index and mobility levels based on Google Community Mobility Reports [6] and Apple mobility trends[7].

Google Community Mobility Reports characterizes movements by classifying mobility into 6 categories. These categories are 1) Retail and Recreation 2) Grocery and Pharmacy 3) Parks 4) Transit stations 5) Workplaces 6) Residential. All the trends are normalized with respect to a baseline from before the pandemic, and hence exhibit primarily negative values. The mobility trends available from Google are on a district level, whereas those from Apple are at the country level. Apple's mobility trends are slightly different in that they directly capture either 'walking' or 'driving' as movement, rather than location based metrics.

Stringency Index encodes the strength of government policies introduced to limit the spread of the Coronavirus. In the context of the Indian subcontinent, a 4-phase lockdown was initially introduced. Post the initial phase, many governments have introduced policies such as night curfews, only essential services, etc. We use the data available on the Our World in Data Platform[8].

For all experiments, we fix the maximum number of **previous observed days to be 28** and number of **forward forecasting days to be 14**. All models are evaluated by forecasting at various times in the past and current year. Training and testing splits are created to avoid biased results. For all models, training segments are created by **non-overlapping sequences** in the dataset. This is to avoid oversampling particular regions of the time series to avoid learning a particular trend.

2) Metrics: We consider the Mean Absolute Percentage Error which is a standard metric used when performing Time Series Forecasting

- Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{1}{n}\sum_{t=0}^{n}|\frac{x_t - \hat{x}_t}{x_t}| \quad (16)$$

We also consider Mean Absolute Error and Root Mean Squared Error

- Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{t=0}^{n}|x_t - \hat{x}_t| \quad (17)$$

- Root Mean Squared Error (RMSE):

$$RMSE = \frac{1}{n}\sum_{t=0}^{n}\sqrt{(x_t - \hat{x}_t)^2} \quad (18)$$

MAPE is a much more reliable and resilient metric than MAE and RMSE as it does not depend on the actual values in the time series. For evaluation, the metrics are computed in multiple different regions in the time series from the test set.

B. Aims

We investigate the following questions in detail:

1) *Can Machine Learning Models perform better or as well as Compartmental Models such as SIR?*

[5]like https://www.mohfw.gov.in/

[6]https://www.google.com/covid19/mobility/

[7]https://covid19.apple.com/mobility

[8]https://ourworldindata.org/grapher/covid-stringency-index

2) *Can Meta Learning be useful in learning cross region patterns so as to allow few shot time series COVID-19 forecasting effectively?*
3) *What role has mobility played in the spread of Coronavirus?*
4) *How effective have government policies been in mitigating the spread of the Coronavirus?*
5) *Have recent events such as election rallies and Kumbh Mela affected the spread of the Coronavirus? In particular, have areas associated with these events been the epicentres of the new spike?*

C. Algorithms and Machines

We utilize Statsmodels [20] for running experiments on Generalized Linear Models, Vector Autoregression and the statistical causality hypothesis tests. We also use Sci-kit Learn [21] for GLMs and for preprocessing data. All experiments are conducted in Python and utilize other standard frameworks such as Numpy [22], Pandas [23] and Matplotlib [24].

We use PyMC3 [25] for Bayesian inference in our SIR model. We implement Deep Learning models in Keras [26] and Pytorch [27]. Most of the Deep Learning experiments are conducted on Google Colaboratory [9]. We utilize the Weights and Biases[10] platform extensively to keep track of experiments and also aid in the analysis of the experiments.

V. Results and Discussion

A. Performance of Machine Learning Models

Performance metrics of all the models on the test set are tabulated in Table II. A few test set forecasts from each model have been shown in Figure 1. Linear autoregressive models perform as good as the commonly used compartmental epidemic forecasting model *SIR* in the metric *RMSE, MAE*. The most commonly used time series forecasting model *ARIMA* suffers as it does not account for exogenous variables such as mobility.

Interestingly, the vanilla variant of the LSTM (with an autoregressive output generation) does not perform well. We attribute to this to the lack of data as we consider only a single timeseries with no overlapping sequences. The vanilla N-beats model is trained for 150 epochs. Our Meta Learning based Transfer Learning variant of N-beats is fine tuned by performing only a single gradient step on a few samples from the training set of the Indian daily COVID counts time series. The performance of the Meta Learning model variant is consistent with the regular N-beats model training. We find the Meta Learning variant of N-beats to be the lowest

It should be noted that by performing training for more epochs, the Meta Learning variant would easily outperform the regular version on all metrics.

[9]https://colab.research.google.com
[10]https://wandb.ai

Model	**MAPE**	RMSE	MAE
Persistence Algorithm	25.22%	6023.71	4795.09
SIR	13.64%	6317.40	5471.85
Autoregressive GLM	14.32%	7688.94	6380.30
Autoregressive LSTM	99.97%	45259.12	42878.34
ARIMA	29.24%	**4702.82**	**3839.01**
Prophet	28.49%	5066.21	4303.10
N-beats	16.90%	8018.86	5755.16
N-beats (with Meta Learning)	**11.53%**	6888.49	6031.47

TABLE II: Performances of the various models

B. Effectiveness of Meta Learning for Few Shot COVID-19 Time Series Forecasting

In this section, we emphasize the effectiveness of Meta Learning as a paradigm for few shot time series forecasting of COVID-19 cases. With the use of Meta-Learning, we show that Deep Learning models can be utilized for forecasting the spread of COVID-19 fairly well with relatively low data samples. Meta Learning model variants may be particularly useful in forecasting the initial stages of disease spread in countries, as the dynamics of the spread of the disease would be similar to others due to the lack of government interventions.

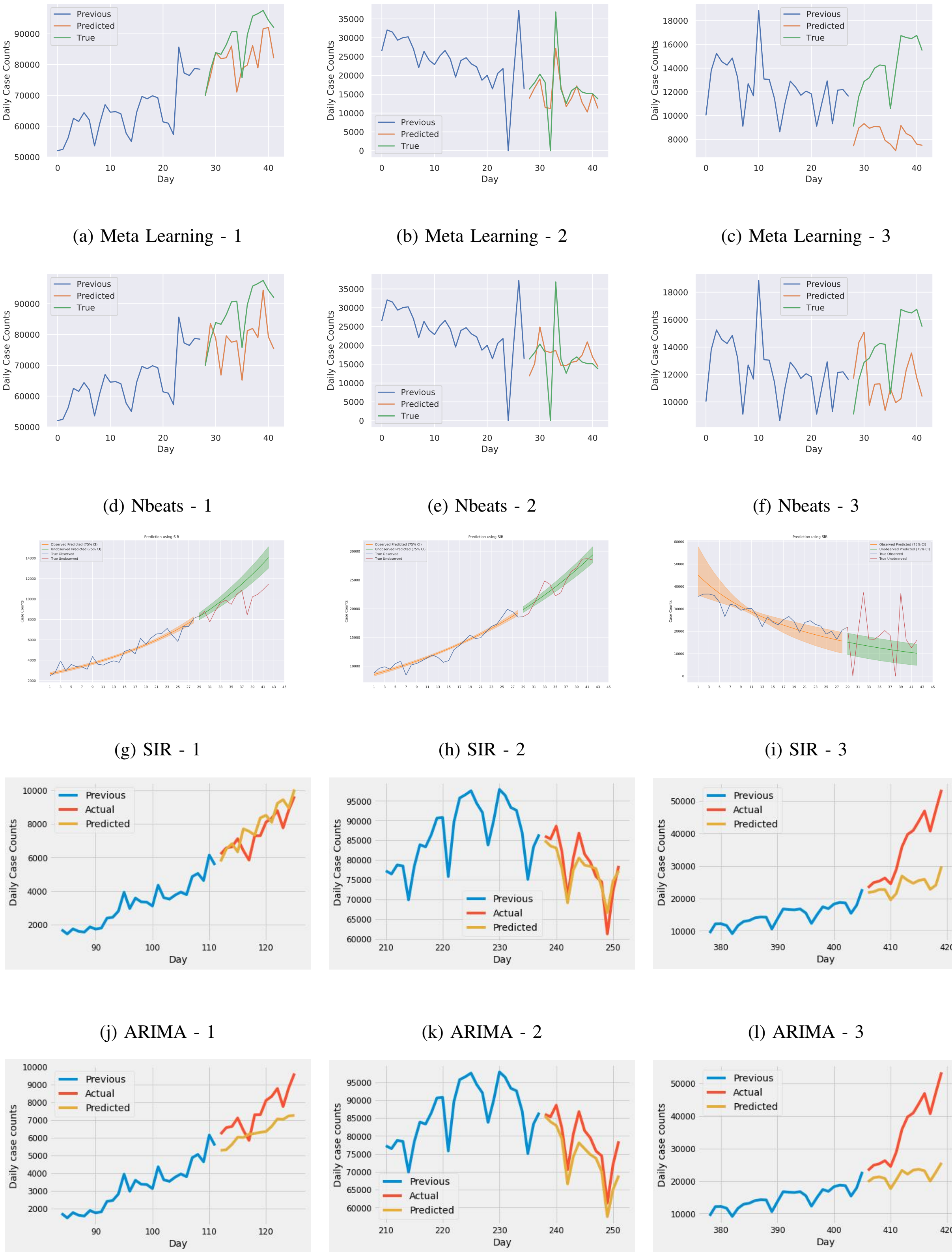
The Meta Learning approach is very scalable in nature as well. Transfer Learning patterns for COVID-19 time series count forecasting could be extracted at various level of demarcated regions i.e., learning common patterns at district level, state level or on a global level. In our current approach we only consider the global scale i.e., how learning about other countries may be helpful in forecasting of our own COVID-19 case counts but our experiments show the results to be consistent at all levels (i.e., district and state level).

The Meta Learning approach also treats each region to be independent i.e., a particular region's case counts may not affect another region's case counts. This is a strong assumption but not entirely valid as the highest risks of disease spread are through travel. Inclusion of region dependent features may boost the performance of the model but it is not clear how directly such information can be encoded.

Finally, the current Meta Learning based approach does not utilize any domain specific knowledge about epidemics and pandemics. That is, it treats the problem as time series forecasting on case counts with a few exogenous features and previous day case counts, without the model being informed about other information such as the knowledge encoded in traditional compartmental models. An approach which accounts for such information, is able to model region dependent features and can also transfer knowledge across countries still remains to be developed.

C. Impact of Lockdown

At the onset of the COVID-19 pandemic, the Indian government implemented a 4-phase nationwide lockdown, restricting the movement of citizens all over the country. In this section,



(a) Meta Learning - 1 (b) Meta Learning - 2 (c) Meta Learning - 3

(d) Nbeats - 1 (e) Nbeats - 2 (f) Nbeats - 3

(g) SIR - 1 (h) SIR - 2 (i) SIR - 3

(j) ARIMA - 1 (k) ARIMA - 2 (l) ARIMA - 3

(m) Prophet - 1 (n) Prophet - 2 (o) Prophet - 3

Fig. 1: Forecasts from all the different models considered

we analyse the mobility patterns of people, and the spread of the virus throughout this period, in order to identify potentially causal relationships.

District	Mobility	p-value	Conclusion
Mumbai	Retail and Recreation	0.259	Accept H_0
	Grocery and Pharmacy	0.019	Reject H_0
	Parks	0.144	Accept H_0
	Transit Stations	0.002	Reject H_0
	Workplaces	0.048	Reject H_0
	Residential	0.945	Accept H_0
Kolkata	Retail and Recreation	0.301	Accept H_0
	Grocery and Pharmacy	0.001	Reject H_0
	Parks	0.359	Accept H_0
	Transit Stations	0.968	Accept H_0
	Workplaces	0.020	Reject H_0
	Residential	0.676	Accept H_0

TABLE III: Results of the statistical F-test on the null hypothesis H_0 that mobility levels do not Granger cause COVID-19 cases

It can be seen in Figure 2 that Phase 1 of the lockdown was relatively effective, with every state reporting highly negative mobility levels (as compared to a baseline). As shown in Figure 2, the spread of the virus also seems to have been under control during this period, with only one state (Maharashtra) reporting more than 100 confirmed cases per day on average.

In the subsequent phases of the lockdown, the graphs in Figure 2 show a gradual increase in mobility levels throughout the country relative to the previous phase. Accompanying this increase in mobility is an increase in confirmed cases of COVID-19. Throughout the lockdown, certain states like Kerala, Andhra Pradesh, Telengana and the eastern states reported relatively low mobility levels as compared to the rest of the country. While their mobility levels did increase from phase to phase, the increase was less pronounced. These states were also particularly effective in containing the spread of COVID-19, being among the states with the lowest daily case counts nationwide. At the opposite end of the spectrum are states like Maharashtra, Tamil Nadu, Gujarat and Uttar Pradesh, which saw significant phase-to-phase increases in mobility, while also reporting relatively high daily case counts.

In summary, a graphical analysis of the mobility patterns and case counts during the lockdown seems to indicate a positive correlation between increased mobility levels and daily case counts. This motivates us to train a model which captures the relationship between mobility levels and daily case counts, and investigate whether there is a causal relationship between the two.

To analyse the relationship between mobility patterns and case counts, we utilize the vector autoregression model mentioned in Section 2 and perform statistical tests for assessing Granger Causality i.e., *does changing mobility affect the occurrence of COVID-19 cases?*. We summarize the results of the statistical tests in Table III.This analysis is done for the entire period (Feb 1, 2020 to April 28, 2021). It can be observed that in major metropolitan cities, we find that Grocercy and Pharmacy Mobility and Workplaces mobility has strong temporal correlation with the case counts. This is expected, as people tend to panic buy and hoard commodities in times of distress, and workplaces are a natural spreading point for the disease, due to the high concentration of people in a confined space. It is also not surprising to see that residential mobility has no linear temporal correlation with the case counts.

D. Effectiveness of Government Policies

In this section, we will establish that the initial lockdown imposed by the government of India was very effective, but since then, government interventional policies have largely played no role in *reducing* the number of cases.

Using a vector autoregression model similar to the one described in Section 2, we performed an F-test with the null hypothesis that the government policies have largely been ineffective or equivalently, that stringency index does not Granger cause COVID-19 cases. Firstly, we choose to analyse this for the initial 90 day period.

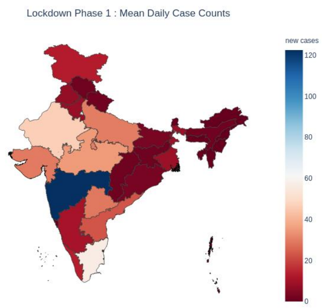
Dates	p-value	Conclusion
Feb 1, 2020 to May 1, 2020	0.024	Reject H_0
May 1, 2020 to August 1, 2020	0.804	Accept H_0
August 1, 2020 to November 1, 2020	0.011	Reject H_0
November 1, 2020 to Feb 1, 2021	0.003	Reject H_0
Feb 1, 2021 to Current	0.845	Accept H_0

TABLE IV: Results of the statistical F-test on the null hypothesis H_0 that government policies such as nation wide lockdown have been ineffective or equivalently, higher stringency index does not Granger Cause lower COVID-19 cases or alternatively, lower stringency index does not Granger Cause higher COVID-19 cases

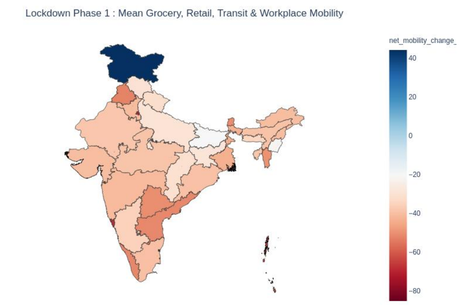
As expected the null hypothesis is rejected for the first period (Feb 1, 2020 to May 1, 2020) to conclude that higher stringency resulted in a slow rise of cases. Interestingly, for the second period (May 1, 2020 to August 1, 2020), we find that there was no effect of stringency in regulating the number of cases. This can be explained by the increasing amount of unrest during Lockdown 3 and 4.

Since the beginning of August, the stringency measures across the country have been largely reduced and hence we see stringency to be a factor in predicting the COVID cases. That is, lower stringency resulted in regular high reporting of cases as compared to the lockdown period (with the first wave peaking in mid-september around 25K cases).

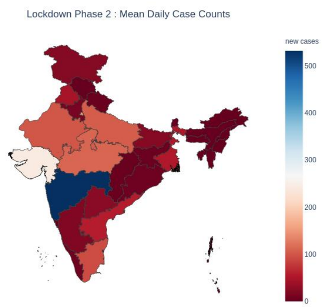
Finally, we observe that after February 1st, there has been no effect of government policies whatsoever with the highest p-value of the linear vector autoregression model to have a zero coefficient i.e., higher stringency does not Granger Cause lower cases. During the months of February, March and April



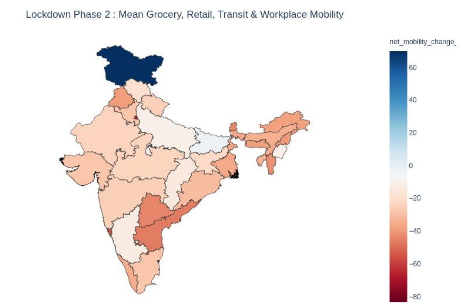
(a) Mean Case Counts during Phase1 of the lockdown.



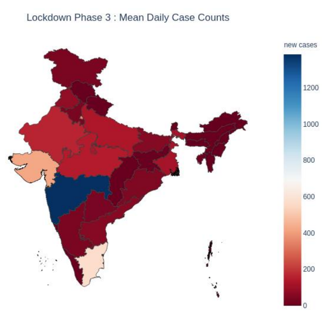
(b) Mean Grocery, Transit, Retail and Workplace Mobility during Phase 1 of the lockdown



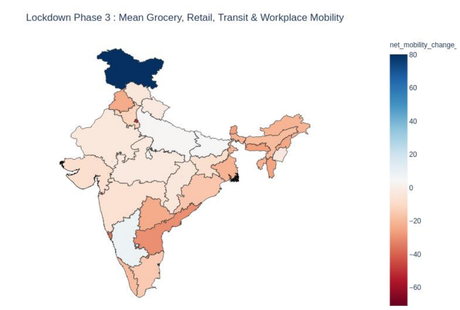
(c) Mean Case Counts during the Phase 2 of the lockdown.



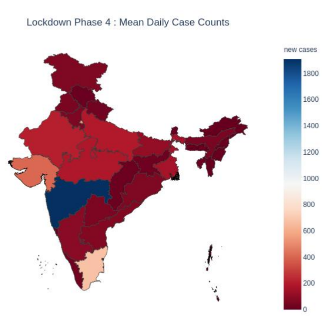
(d) Mean Grocery, Transit, Retail and Workplace Mobility during Phase 2 of lockdown.



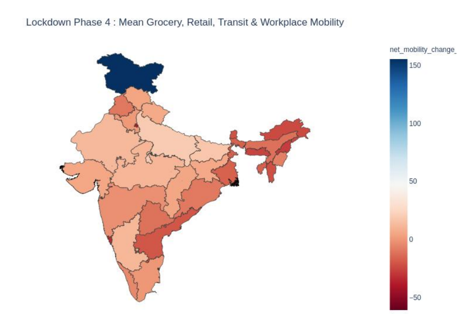
(e) Mean Case Counts during the Phase 3 of the lockdown.



(f) Mean Grocery, Transit, Retail and Workplace Mobility during Phase 3 of lockdown



(g) Mean Case Counts during the Phase 4 of the lockdown.

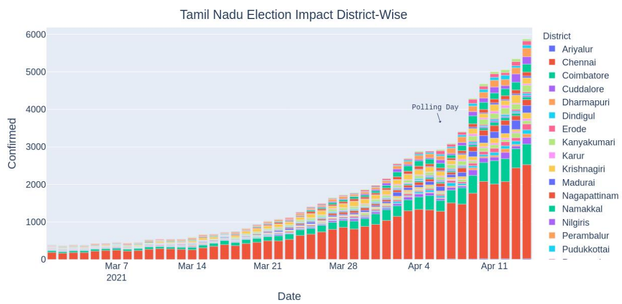


(h) Mean Grocery, Transit, Retail and Workplace Mobility during Phase 4 of lockdown

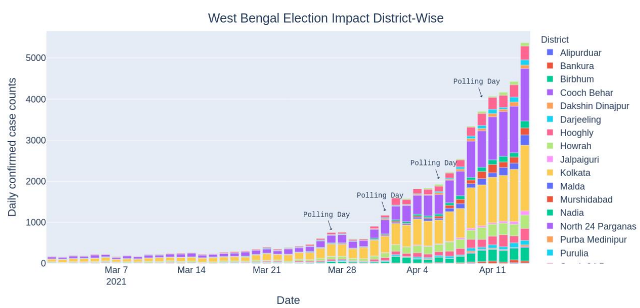
Fig. 2: Case counts and Mobility across the lockdown. Low mobility across all states depicting the effectiveness of the lockdown. Note that while the color scale is the same, the numerical scale varies across graphs. The graph show correlations and hence we analyse if mobility is a cause of Case Counts

2021, we have observed no change in government policy stringency and yet we have witnessed a slow case rise which confirms the results of our analysis.

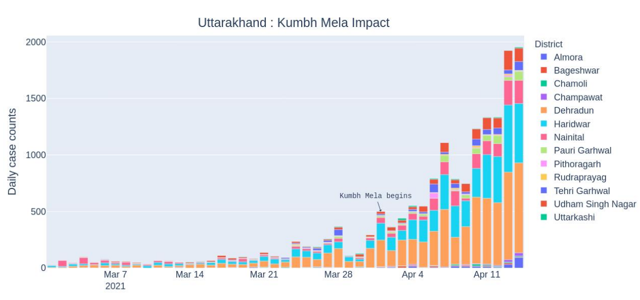
E. Kumbh Mela, Electon Rallies



(a) Impact of Elections Tamil Nadu



(b) Impact of Elections in West Bengal



(c) Kumbh Mela Impact

Fig. 3: Impact of Elections in West Bengal and Tamil Nadu

1) Kumbh Mela: Kumbh Mela is a major festival in Hinduism, which involves large gatherings and communal activities, making it a potential hotspot for the spread of COVID-19. This year, the Kumbh Mela in Haridwar began on April 1, and is scheduled to go on for 1 month. Considering retail and recreation, grocery and pharmacy, transit stations and workplace mobility levels from the GCMR, the net mobility in Uttarakhand leading upto, and during the Kumbh Mela is compared with a baseline in Table V

The net mobility in Uttarakhand doubled during this period, and this coincided with a significant spike in cases during April, as seen in Figure 3c. The Kumbh Mela in 2021 was held in Haridwar. The rise in cases in Uttarakhand in early April had high daily case count reports from two particular regions, Haridwar and Dehradun. This can be explained due to the fact that Dehradun is just north of Haridwar, is the state capital and hence has major transport facilities.

To dive further into this, we analyse if the COVID-19 case counts of Haridwar is a causal feature in some of the districts in Uttarakhand. We pick a vector autoregression model as described in Section III to capture inter-regional correlations between these districts. The null hypothesis for a particular district is that *Haridwar does not Granger Cause that district*. If we reject it at a particular significance level, it means we're concluding Haridwar to Granger cause that district. The results of our analysis are tabulated in Table VI

It can be observed that Haridwar is a strong Granger Causal Feature for most other districts in Uttarakhand. This is confirmed in the Figure 4 which shows that Haridwar experienced the onset of rising cases near the end of March and two weeks later almost every district has reported around the same numbers as Haridwar had reported two weeks before.

Time Period	Net Mobility
Rest of the Year : 14-03-2020 to 23-03-2021	23.635%
Kumbh Mela : 24-03-2021 to 14-04-2021	47.806%

TABLE V: Mobility Levels during the Kumbh Mela and rest of the year

District	p-value	Conclusion
Tehri Garhwal	0.0200	Reject H_0
Champawat	0.0910	Accept H_0
Dehradun	0.0001	Reject H_0
Udham Singh Nagar	0.0001	Reject H_0
Rudraprayag	0.0060	Reject H_0
Pithoragarh	0.0001	Reject H_0
Nainital	0.0090	Reject H_0
Bageshwar	0.0040	Reject H_0
Uttarkashi	0.6660	Accept H_0
Almora	0.4870	Accept H_0
Chamoli	0.4350	Accept H_0

TABLE VI: Granger Causal Analysis of Districts in Uttarakhand. Note: Uttarkashi, Almora and Chamoli report very low cases, hence the outcome is not Granger causal

2) Election Rallies: State elections were held in Tamil Nadu and West Bengal in the months of March and April of 2021. These elections involved extensive campaigning from

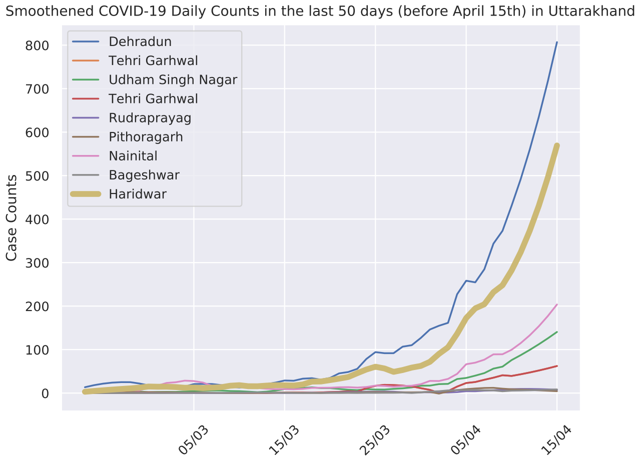


Fig. 4: The Daily Case Counts in Uttarakhand show Haridwar acts as a hotspot of COVID-19 spread. Note: Dehradun is the closest district to Haridwar in Uttarakhand connecting it to the rest of India with transport facilities. Hence, it is a fair assumption that Dehradun also experiences a rise of cases near the end of March due to the Kumbh Mela

State	Mobility Baseline	Election Period Mobility
Tamil Nadu	14.525%	36.953%
West Bengal	-2.607%	17.850%

TABLE VII: Mobility Levels in Tamil Nadu and West Bengal during the election and the previous year

the candidates, including, but not limited to large rallies. A significant spike in cases is seen in Figures 3a and 3b, leading up to the elections in both states.

To judge the impact of campaigning on the mobility patterns of people, we compare the net mobility of each state for the period March 1, 2021 to the last polling day (April 6 for Tamil Nadu and April 14 for West Bengal), with the average mobility over the period March 14, 2020 to February 28, 2021.

As Table VII indicates, the election period in both states saw a significant increase in the average mobility levels of citizens. We also find that the mobility increase primarily is in the major cities of the state, as they serve as transport hubs during large scale rallies. Hence we find that election rallies have had an impact on the case counts in these regions. An analysis similar to one done for Uttarakhand should reveal similar patterns of existence of hotspot regions around places where there have been a lot of election rallies conducted recently.

VI. Conclusion

In this paper, we showed how simple regression models could be used in an autoregressive setting for time series forecasting as well as Compartmental Models. Further, we propose a novel Meta Learning framework to alleviate the issue of less data points for COVID-19 forecasting. Our Meta Learning framework learns in only a single gradient step and performs as well as other models. We also propose to analyse the direct linearly temporal (granger causal) relationship between government policies, mobility and cases counts. We believe our work to be the first in inspiring lots of studies in the domain to analyse the benefits of Meta Learning and also to the use of Causal Inference as a tool in Epidemic Forecasting

Acknowledgment

We would like to thank Prof. Ashwin Srinivasan, Prof. Tirtharaj Dash, and the entire BITS F464 instruction team for their continued assistance and guidance throughout the course of this project.

References

[1] Vivek Shinde, Sutika Bhikha, Zaheer Hoosain, Moherndran Archary, Qasim Bhorat, Lee Fairlie, Umesh Lalloo, Mduduzi S. L. Masilela, Dhayendre Moodley, Sherika Hanley, Leon Fouche, Cheryl Louw, Michele Tameris, Nishanta Singh, Ameena Goga, Keertan Dheda, Coert Grobbelaar, Gertruida Kruger, Nazira Carrim-Ganey, Vicky Baillie, Tulio de Oliveira, Anthonet Lombard Koen, Johan J. Lombaard, Rosie Mngqibisa, As'ad Ebrahim Bhorat, Gabriella Benadé, Natasha Lalloo, Annah Pitsi, Pieter-Louis Vollgraaff, Angelique Luabeya, Aliasgar Esmail, Friedrich G. Petrick, Aylin Oommen Jose, Sharne Foulkes, Khatija Ahmed, Asha Thombrayil, Lou Fries, Shane Cloney-Clark, Mingzhu Zhu, Chijioke Bennett, Gary Albert, Emmanuel Faust, Joyce S. Plested, Andreana Robertson, Susan Neal, Iksung Cho, Greg M. Glenn, Filip Dubovsky, Shabir A. Madhi, and . Preliminary efficacy of the nvx-cov2373 covid-19 vaccine against the b.1.351 variant. *medRxiv*, 2021.

[2] Manindra Agrawal, Madhuri Kanitkar, and Mathukumalli Vidyasagar. Sutra: An approach to modelling pandemics with asymptomatic patients, and applications to covid-19. *arXiv preprint arXiv:2101.09158*, 2021.

[3] I. Rudan. A cascade of causes that led to the covid-19 tragedy in italy and in other european union countries. *Journal of Global Health*, 10, 2020.

[4] I. Rudan. A cascade of causes that led to the COVID-19 tragedy in Italy and in other European Union countries. *J Glob Health*, 10(1):010335, 06 2020.

[5] C. Granger. Investigating causal relations by econometric models and cross-spectral methods. 2001.

[6] Serina Chang, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. Mobility network models of covid-19 explain inequities and inform reopening. *Nature*, 589(7840):82–87, Jan 2021.

[7] Punam Bedi, Shivani, Pushkar Gole, Neha Gupta, and Vinita Jindal. Projections for covid-19 spread in india and its worst affected five states using the modified seird and lstm models, 2020.

[8] Anuradha Tomar and Neeraj Gupta. Prediction for the spread of covid-19 in india and effectiveness of preventive measures. *The Science of the total environment*, 728:138762–138762, Aug 2020. 32334157[pmid].

[9] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, November 1997.

[10] Amol Kapoor, Xue Ben, Luyang Liu, Bryan Perozzi, Matt Barnes, Martin Blais, and Shawn O'Banion. Examining covid-19 forecasting using spatio-temporal graph neural networks, 2020.

[11] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009.

[12] Sarbeswar Praharaj and Hoon Han. A longitudinal study of the impact of human mobility on the incidence of covid-19 in india. *medRxiv*, 2020.

[13] Saleh Albahli and Waleed Albattah. Deep transfer learning for covid-19 prediction: Case study for limited data problems. *Current medical imaging*, 2020.

[14] Yogesh Gautam. Transfer learning for covid-19 cases and deaths forecast using lstm network. *ISA Transactions*, 2021.

[15] Y. Li, W. Jia, J. Wang, J. Guo, Q. Liu, X. Li, G. Xie, and F. Wang. Alert-covid: Attentive lockdown-aware transfer learning for predicting covid-19 pandemics in different countries. *medRxiv*, 2020.

[16] Zhouxuan Li, Tao Xu, Kai Zhang, Hong-Wen Deng, Eric Boerwinkle, and Momiao Xiong. Causal analysis of health interventions and environments for influencing the spread of covid-19 in the united states of america. *Frontiers in Applied Mathematics and Statistics*, 6:69, 2021.
[17] Kristi Bushman, Konstantinos Pelechrinis, and Alexandros Labrinidis. Effectiveness and compliance to social distancing during covid-19. *arXiv preprint arXiv:2006.12720*, 2020.
[18] Sean J Taylor and Benjamin Letham. Forecasting at scale. September 2017.
[19] Boris N. Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-beats: Neural basis expansion analysis for interpretable time series forecasting, 2020.
[20] Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.
[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
[22] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.
[23] Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010.
[24] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.
[25] John Salvatier, Thomas V. Wiecki, and Christopher Fonnesbeck. Probabilistic programming in python using PyMC3. *PeerJ Computer Science*, 2:e55, April 2016.
[26] Francois Chollet et al. Keras, 2015.
[27] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.