Development of a spatial and temporal based COVID-19 predictor for Pakistan

Coronavirus (COVID-19) has expanded across Pakistan in a large-scale outbreak with peculiar characteristics [6]. The importance of a strong national coordinating level for consistent implementation of control measures, as well as the value of local forecasts, are highlighted by differences in epidemic dynamics in Pakistan. One of the most often requested issues by public officials since the breakout of COVID-19 has been how to determine the peak of this dangerous disease. Analyses and forecasts in both space and time can show which regions will be impacted the most and when [11]. This type of analysis can provide policymakers adequate time to intervene in local policy.

Geographical Spread Pattern of COVID-19

COVID-19 virus spreads spatially and temporally [1, 2, 3, 4]. Geographical-temporal spread patterns were discovered utilising approaches such as geographic clustering, hotspot identification, and spatial heterogeneity and homogeneity techniques [1], and the direction of virus transmission was determined using spatial heterogeneity and homogeneity techniques [1]. In addition, techniques including geographic statistical analysis, epicentre incidence rates, and the "space-specific strategy" are used to identify ways for limiting the virus's spatial spread [1].

According to Tobler's first law, [1, 4] it is primarily concerned with where the disease is concentrating and spreading. in order to figure out how a disease spreads through time and place. In all research based on COVID-19 spatial-temporal analysis, the maps of respective nations are coloured on a specified scale on a weekly/monthly basis to assess the virus's propagation, and different factors such as hotspots, epicentres, and cold spots are depicted [1, 2].

The transmission of the COVID-19 virus was found to be influenced by the following geographical conditions in previous studies.

- Previous viruses such as influenza, pneumonia, respiratory Syncytial virus, malaria, dengue, zika, HIV, and Choles had the similar geographical pattern (high probability of a spatial pattern) [1, 2].
- According to the study [1,] the transmission of infectious diseases, particularly COVID-19, is also linked to human migration.
- Infectious illnesses are geographically specific [1, 2].

Factors Contributing in the Spread of COVID-19

COVID-19 instances and temperature had a positive association, according to the correlation analysis. It means that a spike in COVID-19 cases was observed in Pakistan, its provinces, and administrative units, as a result of rising temperatures. Second, in Pakistan, there was no link between rainfall and COVID-19 instances. Finally, there was a negative association between humidity and total COVID-19 instances, implying that increasing humidity is advantageous in preventing COVID-19 transmission across Pakistan, its provinces, and administrative units. In Sindh province, humidity and COVID-19 instances were found to be positively associated [15].

COVID-19 instances are influenced by a variety of factors, including geographical landscapes, economic situations, demographic illness variations, genetic factors, health care system, number of tests, and age disparities, among others. [5] A huge number of undetected infected persons is also a crucial factor contributing to the spread of COVID-19 [9].

Methodologies used to predict the Spread of COVID-19

Methodologies like the ARIMA model, which is solely focused on predicting epidemiological variables like cumulative cases, deaths, and recoveries from the current COVID-19 epidemic. As a result, ARIMA (1, 0, 4) was chosen as the best ARIMA model for evaluating COVID-19 prevalence, whereas ARIMA (1, 0, 3) was recommended for COVID-19 prediction [6].

Time series models are utilised to capture the trends and patterns of all infectious disease-related events, particularly infectious illnesses with cyclical patterns like influenza [6]. As a result, similar methodologies have been utilised to uncover patterns in COVID-19 occurrences since it replicates a geographic pattern similar to that identified in earlier infectious viruses [1, 2].

Single exponential smoothing, Holt linear trend technique, Holt winter method, and the auto-regressive integrated moving average (ARIMA) model, which was originally created for economics applications, are among the forecasting methods utilised. Two statistical indices, RMSE (Root Mean Square Error) and MAE (Mean Absolute Error), are used in the above-mentioned models to determine the best candidate model for predicting [6].

The temporally dependent To fit the data and make future predictions, the Susceptible-Infected-Recovered (SIR) model is used. The day when the transmission rate falls below the recovery rate is considered the turning point of the pandemic [7].

The SIR model is based on two variables that can be changed;

- $\beta(t)$, which measures the transmissions per unit of time
- $\gamma(t)$ that measures the recoveries per unit time.

These characteristics can be accurately estimated to forecast a realistic number of illnesses and the pandemic's peak. To estimate their values from data and predict future trends, the exponential fit is used [7].

Using previous data, an artificial neural network with a rectifying linear unit-based technique is used to estimate the number of deaths, recovered, and confirmed cases of COVID-19 in Pakistan. The acquired data was separated into training and test data, which were used to evaluate the suggested technique's effectiveness. An ANN is made up of neurons, which are small units that can transmit signals to other neurons, and numerous neurons working together to create a prediction [8].

The multivariate Long Short-term Memory (LSTM) recurrent neural network, which includes a mobility series, is trained on many time series samples at the same time. The results suggest that include mobility as a variable and training the network with many samples improves prediction performance in terms of bias and variance. Many features of the model appeal to me, notably the fine spatial granularity of predictions and excellent predictive performance several weeks ahead. Furthermore, substituting a single model for several models folded in an ensemble model reduces data requirements and computing effort [11].

Although not the most efficient method, employing a GIS environment has some advantages, including access to a wide range of geographical functions and the ability to design and prototype quickly. This technique, on the other hand, is helpful for quickly tracing and treating patients, as well as implementing preventative steps in that specific location to stop COVID-19 from spreading. As a result, local authorities must isolate that region rather than the entire country [13]. In addition, the GIS technique may be utilised to provide a comprehensive picture of the illness source, dynamics and response mapping, risk assessment, and creating quarantine areas to keep COVID-19 patients out [14].

We turn places into nodes on a network and generate edges based on geographical proximity and demographic similarity in the Spatio-Temporal Attention Network (STAN) for pandemic prediction based on real-world evidence. Each node is linked to a set of static and dynamic features derived from a range of real-world evidence in medical claims data, such as illness prevalence and medical resource consumption conditions. We employ a graph attention network to connect the interactions of similar sites. Then, utilising epidemiological models' transmission dynamics, we forecast the number of infected people for a given time period while simultaneously setting physical restrictions on forecasts. When compared to the best baseline model [10], we achieve a mean squared error reduction of up to 87 percent when using STAN to anticipate future numbers of infected patients at the state and county levels.

Motivation for using Machine Learning

One of the main motivations for using machine learning to predict the COVID-19 pandemic is that it can more accurately estimate the effects of social distancing measures on virus transmission while taking into account socio-economic control factors, which are difficult to capture using traditional econometric methods or compartmental models. To construct a universal, dependable, and trustworthy model that provides consistent predictions across varied systems (e.g., data, settlement situations), high-quality data free of collecting biases, robust techniques, and validation with external data sets and models are required [11].

Challenges faced by Epidemiological models and Deep learning models

1. They frequently develop separate models for each area (for example, one model per county) without taking into account physical proximity or relationships with neighbouring regions. Alternatively, projections may be based only on trends observed in other regions [10].
2. The majority of existing models are based on COVID-19 case report data. There is a lot of underreporting in these data, as well as other data quality concerns [10].

3. There are deterministic models available, such as SIR and SEIR. They use a set of differential equations to fit the entire curve of disease numbers. These models are unable to capture intricate short-term patterns such as superinfection or time-varying infectivity since they are dependent on a small set of variables [10].
4. Deep learning-based models, on the other hand, can only predict known data patterns and can only make correct predictions in a limited amount of time [10].

Conclusion

Researchers frequently utilise spatio-temporal models to describe the temporal progression of disease across geographic regions, as well as patterns of infection and mortality rates. Furthermore, when compared to classic epidemiological and machine learning-based models, spatio-temporal models produce better predictions, with higher prediction accuracy and a lesser tendency to over- or under-fit [11]. As a result, adopting spatial-temporal-based predictive models at the province level can help public decision-makers better plan health-policy activities by significantly improving projections of sick people.

References

1. Bag, Rakhohori, et al. "Understanding the spatio-temporal pattern of COVID-19 outbreak in India using GIS and India's response in managing the pandemic." *Regional Science Policy & Practice* 12.6 (2020): 1063-1103.

2. Hazbavi, Zeinab, et al. "Spatial and temporal analysis of the COVID-19 incidence pattern in Iran." *Environmental Science and Pollution Research* 28.11 (2021): 13605-13615.

3. Huang, Rui, Miao Liu, and Yongmei Ding. "Spatial-temporal distribution of COVID-19 in China and its prediction: A data-driven modeling analysis." *The Journal of Infection in Developing Countries* 14.03 (2020): 246-253.

4. Sartorius, B., A. B. Lawson, and R. L. Pullan. "Modelling and predicting the spatio-temporal spread of COVID-19, associated deaths and impact of key risk factors in England." *Scientific reports* 11.1 (2021): 1-11.

5. Mehmood, Khalid, et al. "Spatiotemporal variability of COVID-19 pandemic in relation to air pollution, climate and socioeconomic factors in Pakistan." *Chemosphere* 271 (2021): 129584.

6. Ali, Muhammad, et al. "Forecasting COVID-19 in Pakistan." *Plos one* 15.11 (2020): e0242762.

7. Waqas, Muhammad, et al. "Analysis and prediction of COVID-19 pandemic in Pakistan using time-dependent SIR model." *arXiv preprint arXiv:2005.02353* (2020).

8. Ahmad, Iftikhar, and Syed Muhammad Asad. "Predictions of coronavirus COVID-19 distinct cases in Pakistan through an artificial neural network." *Epidemiology & Infection* 148 (2020).

9. Shwartz-Ziv, Ravid, Itamar Ben Ari, and Amitai Armon. "Spatial-Temporal Convolutional Network for Spread Prediction of COVID-19." *arXiv preprint arXiv:2101.05304* (2020).

10. Gao, Junyi, et al. "STAN: spatio-temporal attention network for pandemic prediction using real-world evidence." *Journal of the American Medical Informatics Association* 28.4 (2021): 733-743.

11. Nikparvar, Behnam, et al. "Spatio-temporal prediction of the COVID-19 pandemic in US counties: modeling with a deep LSTM neural network." *Scientific reports* 11.1 (2021): 1-12.

12. Vahedi, Behzad, Morteza Karimzadeh, and Hamidreza Zoraghein. "Spatiotemporal prediction of COVID-19 cases using inter-and intra-county proxies of human interactions." *Nature communications* 12.1 (2021): 1-15.

13. Sarwar, Suleman, et al. "COVID-19 challenges to Pakistan: Is GIS analysis useful to draw solutions?." *Science of the Total Environment* 730 (2020): 139089.

14. Saeed, Urooj, et al. "Identification of potential lockdown areas during COVID-19 transmission in Punjab, Pakistan." *Public health* 190 (2021): 42-51.

15. Raza, Ali, et al. "Association between meteorological indicators and COVID-19 pandemic in Pakistan." *Environmental Science and Pollution Research* 28.30 (2021): 40378-40393.

16. Khakharia, Aman, et al. "Outbreak prediction of COVID-19 for dense and populated countries using machine learning." *Annals of Data Science* 8.1 (2021): 1-19.