**Literature Review**

Our research found an abundance of articles that used time series models to forecast the spread of covid-19. We saw that from the start, in March 2020, when the World Health Organization (“WHO”) declared COVID-19 a global pandemic, researchers deployed an array of time series models to forecast the spread of the virus. These efforts were largely done to inform the general public of the need to enact mitigation measures to stop the spread. The articles discussed in this review provide an exemplar of the articles that we researched. We begin our review by comparing the findings from very basic time series models and finishing with a look at the more advanced ones.

Authors Haytham H. Elmousalami and Aboul Ella Hassanien’s research focused on the day level spread of the virus [1]. Their methodology used naive time series models, Moving Average (“MA”), Weighted Moving Average (“WMA”), which take averages or weighted averages of past observations to forecast future cases. Such models are easy to create and even easier to report to the public. Most media outlets include a moving average component when reporting about the spread. The other model that the authors used was Single Exponential Smoothing (“SES”), another naïve model. In this instance, instead of weighing past observations equally, SES uses functions to exponentially decrease the weighting of past observations.

The results of Elmousalami and Hassanien’s models showed that the SES model had the highest accuracy for confirmed cases, recovered cases, and deaths based on the evaluations of Mean Absolute Deviation (“MAD”), Mean Square Error (“MSE), Root Mean Square Error (“RMSE”), and Mean Absolute Percentage Error (“MAPE”). In contrast to Elmousalami and Hassanien, authors Vasilis Papastefanopoulos, Pantelis Linardatos and Sotiris Kotsiantis used a higher class of time series models [2], Auto Regressive Integrated Moving Average (“ARIMA”), Holt-Winters additive model (“HWAM”), Trigonometric seasonal formulation Box-Cox transformation ARIMA errors and trend component (“TBAT”), Facebook’s Prophet, Deep AR, a probabilistic forecasting with Auto-Regressive Recurrent Networks, and N-Beats, a neural basis expansion analysis for interpretable time series forecasting.

Papastefanopoulos, Linardatos, and Kotsiantis’ models consisted of linear regression and deep learning neural networks. Instead of applying averages or weights or exponentially decreasing weights, these models primarily make predictions by using either a regression of past observations or by using a system of inter-connected nodes that learns from past observations. Also, in contrast to Elmousalami and Hassanien, the authors Papastefanopoulos, Linardatos, and Kotsiantis did not find a “one-size-fits-all” model. Their findings showed that based, on RMSE measures, the ARIMA and TBAT models performed best in most of the countries while achieving second best in the other two. They found “traditional statistical methods such as such ARIMA and TBAT overall prevail over deep learning counterparts such as DeepAR, and N-BEATS—an outcome which, **due to the lack of large amounts of data** [2, p. 11].” (emphasis added)

Authors Vinay Kumar, Reddy Chimmula, and Lei Zhang focused solely on using a single deep learning network, Long short term Memory (“LSTM”) [3], a non-linear approach that uses a Recurrent Neural Network (“RNN”) to forecast trends. In an RNN, output from the last step is fed as input to the current step. This is somewhat similar to Elmousalami and Hassanien’s approach where the weights applied to past observations are manipulated to make forecasts. The difference is that in LSTM networks can retain long term information which is useful if there are lags of unknown duration between important time gaps.

Using data collected in Canada until March 31, 2020, Kumar, Chimmula, and Zhang’s methodology was to use sequential networks to extract the patterns from a time series dataset. The rationale for this approach was that the linear approach often neglects the temporal com- ponents in the data. “They depend upon regression without non- linear functions and failed to capture the dynamics of transmission of infectious diseases like novel corona virus. Statistical models such as Auto Regressive Integrated Moving Average (ARIMA), Moving Average (MA), Auto Regressive (AR) methods overwhelmingly depends on assumptions and such models are difficult for forecasting real-time transmission rates.” [3, p. 1] In contrast to Papastefanopoulos, Linardatos, and Kotsiantis, they showed that the RMSE of the LSTM had the highest accuracy.

Authors Abdelhafid Zeroual et al. [4] differed from the previously discussed studies by comparing the five most advanced models, Recurrent Neural Network (“RNN”), Long short-term memory (“LSTM”), Bi-directional LSTM (“Bi-LSTM”), Gated recurrent units (“GRUs”) and Variational AutoEncoder (“VAE”) to forecast cases and recovered cases across six countries. The authors cited the models’ ability to “handling temporal dependencies in time series data, distribution-free learning models, and their flexibility in modeling nonlinear features.” [4, p.2]

Using RMSE as their primary performance metric, the authors found that VAE outperformed the other models for confirmed and recovered cases. This study was one of the first times that VAE has been used to model COVID-19 cases. The authors offer a reason as to why VAE out-performed the other advanced models. “[T]he capacity of the VAE in dealing with small data compared to the other recurrent models (RNN, L STM, Bi-L STM, and GRU) **which may need more lengthy data to extract relevant variability in time series data** [4, p. 10].” (emphasis added) Poor performance of advanced models due to the lack of data was the same issue that Papastefanopoulos, Linardatos, and Kotsiantise experienced with the performance of DeepAR and N-BEATS. However, we did not see this issue with Kumar, Chimmula, and Zhang’s LSTM model.

The articles discussed and compared in our review are indicative of the articles that we researched for the project. Each one chose a particular class of models to compare, used various datasets, compared the results using standard metrics, and summarized their findings. What we did not find was one paper that compared all classes of models, from the naive to the more advanced, in one study over one single dataset over the same period of time in order to determine which model(s) performed the best. That is the objective of this paper.

**Key Words**

*Pandemic, Covid, Time Series, MAPE, RMSE, ARIMA, Prophet, Cases, Predictions, SARIMA, Forecasting, Comparison, Long short-term Memory (“LSTM”), Recurrent Neural Network (“RNN”), Bidirectional LSTM (“BiLSTM”), Gated Recurrent Units (“GRUs”), and Variational AutoEncoder (“VAE”)*

**References**

1. **Haytham H. Elmousalami and Aboul Ella Hassanien, “Day Level Forecasting for Coronavirus Disease (COVID-19) Spread: Analysis, Modeling and Recommendations”, *Scientific Research Group in Egypt (SRGE), Cairo, Egypt* (March 15, 2020)**
2. **Papastefanopoulos, Vasilis, Linardatos, Pantelis, and Kotsiantis, Sotiris, “COVID-19: A Comparison of Time Series Methods to Forecast Percentage of Active Cases per Population” *Applied Sciences. 10. 3880. 10.3390/app10113880* (May 2020)**
3. **Kumar,Vinay, Chimmula, Reddy, and Zhang ,Lei “Time series forecasting of COVID-19 transmission in Canada using LSTM networks****”, *Chaos, Solitons and Fractals Nonlinear Science, and Nonequilibrium and Complex Phenomena* (May 2020)**.
4. **Zeroual, Abdelhafid, Harrouc ,Fouzi, Dairi, Abdelkader, and Sunc, Ying, “Deep learning methods for forecasting COVID-19 time-Series data: a Comparative study”, *Chaos, Solitons and Fractals Nonlinear Science, and Nonequilibrium and Complex Phenomena* (July 15, 2020**)
5. **Kumar, Naresh & Susan, Seba, “COVID-19 Pandemic Prediction using Time Series Forecasting Models”, *The 11th ICCCNT 2020 conference*.**
6. **Er, Başak, Emeç, Murat, and Ozcanhan, Mehmet, “Analysis of COVID-19 Data Using Arima Time Series Model”, *Conference: V. International Scientific And Vocational Studies Congress – Engineering* (December 2020).**
7. **Mahmud, Sakib, Bangladesh COVID-19 Daily Cases Time Series Analysis using Facebook Prophet Model, *Social Science Research Network***
8. **Ismail, Khan, Znati, Materwala, Turaev, ”Tailoring time series models for forecasting coronavirus spread: Case studies of 187 countries”, *Computational and Structural Biotechnology Journal Volume 18, 2020, Pages 2972-3206* (September 2020)**.

1. **Chaurasia, Vikas and Pal,Saurabh, “Application of machine learning time series analysis for prediction COVID-19 pandemic”, *Sociedade Brasileira de Engenharia Biomedica* (October 2020)**.
2. **Vijander Singh , Ramesh Chandra Poonia , Sandeep Kumar, Pranav Dass , Pankaj Agarwal , Vaibhav Bhatnagar & Linesh Raja, “Prediction of COVID-19 corona virus pandemic based on time series data using support vector machine”, *Journal of Discrete Mathematical Sciences and Cryptography*, 23:8, 1583-1597, DOI: 1080/09720529.2020.1784535.**
3. **Shah, Saloni; Mulahuwaish, Aos; Ghafoor, Kayhan; Maghdid, Halgurd S., “Prediction of Global Spread of Covid-19 Pandemic: A Review and Research Challenges.” *TechRxiv. Preprint*.** [***https://doi.org/10.36227/techrxiv.12824378.v1***](https://doi.org/10.36227/techrxiv.12824378.v1)
4. **Yi-Cheng Chen, Ping-En Lu , Cheng-Shang Chang, “A Time-Dependent SIR Model for COVID-19 with Undetectable Infected Persons”, *IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, VOL. 7, NO. 4*, (OCTOBER-DECEMBER 2020)**
5. **Ian Cooper, Argha Mondal , Chris G. Antonopoulos, “A SIR model assumption for the spread of COVID-19 in different communities”, *Chaos, Solitons and Fractals 139* (2020)**
6. **Vinay Kumar, Reddy Chimmula, and Lei Zhan, “Time series forecasting of COVID-19 transmission in Canada using LSTM networks”, Chaos, Solitons and Fractals (May 2020)**