Covid - 19 Forecasting & Prediction Using Machine Learning

Prof. Amit Thakur[1] Sourabh Sahu[2] Dikendra Kumar[3] A Ashutosh[4] Sheikh Shoeb[5]

[1,2,3,4,5]Department of Computer Science & Engineering

[1,2,3,4,5]Bhilai Institute of Technology, Raipur, India

Abstract— Corona virus or COVID-19 may be a hazardous disease that has endangered the health of many people around the world by directly affecting the lungs. COVID-19 is a medium-sized, coated virus with a single-stranded RNA. This virus has one of the largest RNA (Ribonucleic Acid) genomes and is approximately 120 nm. The X-Ray and computerized tomography (CT) imaging modalities are widely wont to obtain a quick and accurate diagnosis. Identifying COVID-19 from these medical images is extremely challenging because it is time-consuming, demanding, and. During this paper, a complete survey of studies on the appliance of ML techniques for COVID-19 diagnostic and automatic segmentation of lungs is discussed in this field; we are concentrating on works. When a review of papers on the forecasting of corona virus prevalence in several parts of the planet with ML techniques is presented. Lastly, the challenges faced within the automated detection of COVID-19 using ML techniques and directions for future research are discussed. Anticipated the potential patterns of COVID-19 effects in India as well as in the world dependent on data gathered from Kaggle.

Keywords: Covid Forecasting, genomes, RNA, CT, X-Ray, Machine Learning

I. INTRODUCTION

In December 2019 the novel COVID-19 or corona virus came to light in Wuhan Province, China, where it originated from animals of the seafood market and quickly spread around the world. The easiest way to transmit COVID-19 is through physical contacts, such as hand contact with infected people [1]. The virus inserts itself into the lung cells through the respiratory system and replicates there, by destroying these cells [2].

Every country imposed different methodologies or strategies starting from staying in-home, using masks, travel restrictions, avoiding social gatherings, avoiding hand sacking, frequently washing hands and sanitizing the places often in the case of a common effort to combat the outbreak of this disease. Many countries imposed a lockdown state that prevents the movement of the citizens unnecessarily [3]. GPD (Growth) of the entire world dropped drastically. When the person is found infected from covid-19, he is isolated and treatment is given for recovery.

There is a key to understanding the pandemic starts with an understanding of the disease itself, and the progression of the natural course of the disease. The disease is often defined as the state that negatively affects the body of a living person, plant or animal. The disease affects the body because of a pathogenic infection. The natural course of the disease starts before the onset of the infection. After which it progresses through the pre-symptomatic stage. The last stage is the clinical phase. In the clinical phase, a patient gets the prognosis of the disease. After a successful treatment of the disease, the patient enters into the remission stage. Remission refers to a decrease in the symptoms or a complete disappearance of the disease. The patient needs to follow instructions given by the doctor very strictly in the remission stage. This will ensure that the disease does not recur. If treatment is not successful a patient can be dead or chronically disabled [4].

There are some important terms used to represent the statistics:

1) Case-fatality rate: It is defined as the ratio of the number of patients who die due to the disease to the number of people who have it.
2) Observed survival rate: it is the prediction of the probability of survival.
3) Relative survival rate: It is defined as the ratio of the observed survival to the expected survival.

Fig. 1: Host formation & progression

II. CORONAVIRUS OVERVIEW

A Corona virus is a kind of common virus that causes an infection in your nose, sinuses, or upper throat. Most corona viruses aren't dangerous. In early 2020, after a December 2019 outbreak in China, the World Health Organization identified SARS-CoV-2 as a new type of corona virus. The outbreak quickly spread around the world. COVID-19 is a disease caused by SARS-CoV-2 that can trigger what doctors call a respiratory tract infection. It can affect your upper respiratory tract (sinuses, nose, and throat) or lower respiratory tract (windpipe and lungs). It spreads the same way other corona viruses do, mainly through person-to-person contact. Infections range from mild to deadly. SARS-CoV-2 is one of seven types of corona virus, including the ones that cause severe diseases like the Middle East respiratory syndrome (MERS) and sudden acute respiratory syndrome (SARS). The other corona viruses cause most of the colds that affect us during the year but aren't a serious threat for otherwise healthy people.

In the case of SARS-CoV-2, an efficacious vaccine might prevent infection, disease, or transmission. The outcome of SARS-CoV-2 infection in individuals is heterogeneous and dependent on multiple variables, including age, sex, ethnicity, and co-morbidities. On an individual level, the consequence of infection can range from paucisymptomatic states to hospital admission, requirement for respiratory support, and death. Transmission dynamics of SARS-CoV-2 are not yet fully understood but the ability of infected individuals to transmit infection when asymptomatic or in a presymptomatic period means that infection control strategies that focus solely on preventing transmission from symptomatic individuals will be insufficient alone to interrupt the transmission of SARS-CoV-2.

The effect of an efficacious vaccine on the course of the SARS-CoV-2 pandemic is complex and there are many potential scenarios after deployment. The ability of a vaccine to protect against severe disease and mortality is the most important efficacy endpoint, as hospital and critical-care admissions place the greatest burden on healthcare systems. However, the beneficial effects of such a vaccine on a population can be observed only if the vaccine is efficacious in older adults (e.g. approximately >60 years) and widespread distribution of the vaccine exists, including to people who are most susceptible to COVID-19. High coverage among these groups who are at high risk of severe COVID-19 would have the greatest effect against disease endpoints. Alternatively, vaccines that do not affect the clinical course, but reduce the transmissibility of SARS-CoV-2, could still be valuable interventions on a population level.

III. LITERATURE REVIEW

1) R. Sujatha, Jyotir Moy Chatterjee and Aboul Ella Hassanien (2020) proposed a model to predict the spread of COVID-2019 by using linear regression, Multilayer perceptron and Vector autoregression model on the COVID-19 Kaggle data to envision the epidemiological example of the malady and pace of COVID-2019 cases in India.
2) Gitanjali R. Shinde, Asmita B. Kalamkar, Parikshit N. Mahalle, Nilanjan Dey, Jyotismita Chaki and Aboul Ella Hassanien (2020) present two categories of datasets that have been discussed i.e. big data accessed from WHO/National databases and data from a social media communication. Forecasting of a pandemic can be done based on various parameters such as the impact of environmental factors, incubation period, the impact of quarantine, age, gender and many more. These techniques and parameters used for forecasting are extensively studied in this work. However, forecasting techniques come with their own set of challenges (technical and generic). This study discusses these challenges and also provides a set of recommendations for the people and the governments that are currently fighting the global COVID-19 pandemic.
3) Afshin Shoeibi, Marjane Khodatars, Roohallah Alizadehsani, Navid Ghassemi, Mahboobeh Jafari, Parisa Moridian, Ali Khadem, Delaram Sadeghi, Sadiq Hussain, Assef Zare, Zahra Alizadeh Sani, Javad Bazelim, Fahime Khozeimeh, Abbas Khosravi, Saeid Nahavandi, U. Rajendra Acharya, Senior Member, and Peng Shi (2020) created a complete survey of studies on the application of DL techniques for COVID-19 diagnostic and automated segmentation of lungs is discussed, concentrating on works that used X-Ray and CT images and a review of papers on the forecasting of coronavirus prevalence in different parts of the world with DL techniques is presented. Lastly, the challenges faced in the automated detection of COVID-19 using DL.
4) Dianbo Liu, Leonardo Clemente, Canelle Poirie, Xiyu Ding, Matteo Chinazzi, Jessica Davis, Alessandro Vespignani, Mauricio Santillana proposed a model to produce stable and accurate forecasts 2 days ahead of the current time and outperforms a collection of baseline models in 27 out of 32 Chinese provinces.
5) Cleo Anastassopoulou, Lucia Russo, Athanasios Tsakris, Constantinos Siettos provide an estimation of the case fatality and case recovery ratios, along with their 90% confidence intervals as the outbreak evolves. On the basis of a Susceptible-Infectious-Recovered-Dead (SIDR) model, we provide estimations of the basic reproduction number ($R0$), and the per day infection mortality and recovery rates. By calibrating the parameters of the SIRD model to the reported data.
6) K.C. Santosh discussed how continuous and unprecedented factors lead us to design complex models, rather than just relying on stochastic and/or discrete ones that are driven by randomly generated parameters. Further, it is a time to employ data-driven mathematically proven models that have the luxury to dynamically and automatically tune parameters over time.
7) Dianbo Liuy, Leonardo Clementey, Canelle Poiriery, Xiyu Ding, Matteo Chinazzi, Jessica T Davis, Alessandro Vespignani, Mauricio Santillana presented a timely and novel methodology that combines disease estimates from mechanistic models with digital traces, via interpretable machine-learning methodologies, to reliably forecast COVID-19 activity in Chinese provinces in real-time. Specifically, the method is able to produce stable and accurate forecasts 2 days ahead of the current time, and uses as inputs (a) official health reports from Chinese Center Disease for Control and Prevention (China CDC), (b) COVID-19-related internet search activity from Baidu, (c) news media activity reported by Media Cloud, and (d) daily forecasts of COVID-19 activity from GLEAM, an agent-based mechanistic model. Our machine-learning methodology uses a clustering technique that enables the exploitation of geo-spatial synchronicities of COVID-19 activity across Chinese provinces, and a data augmentation technique to deal with the small number of historical disease activity observations, characteristic of emerging outbreaks.

The University of Pennsylvania named their prediction model, CHIME: COVID-19 Hospital Impact Model for Epidemics. Their model allows users to vary inputs and assumptions. In their predictions, for the next three months, they forecasted best- and worst-case scenarios for a total number of hospitalizations, ICU bed demand, ventilator demand, and a number of days these demands would exceed hospitals capacities.

IV. FORECASTING TECHNIQUES

In this paper, forecasting has been done based on fbprophet techniques and different data sources. To understand and improve the forecasting this section categorizes these techniques into multiple types for better analysis. This categorization is done based on the data sources used i.e. Kaggle dataset for the latest updates regarding active cases, deaths and recovered patients. Categorization is also done based on fbprophet techniques that are used for forecasting i.e. machine learning techniques.

Various statistical, analytical, mathematical and medical (symptomatic and asymptomatic) parameters are taken into consideration for analysis. However, major significant parameters are listed below:

1) Death count per day

2) Number of carriers
3) Incubation period
4) Awareness about covid – 19
5) Social distancing, quarantine, isolation
6) Transmission rate
7) Age and gender
8) Government policies and many

A. Data Collection (Kaggle Dataset)

For data collection, we used Kaggle dataset in our paper. Kaggle, is a subsidiary of Google LLC, is an online community of data scientists and machine learning practitioners. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges.

Kaggle got its start in 2010 by offering machine learning competitions and now also offers a public data platform, a cloud-based workbench for data science, and Artificial Intelligence education. Its key personnel were Anthony Goldbloom and Jeremy Howard. Nicholas Gruen was founding chair succeeded by Max Levchin.

B. Data Visualization (choropleth)

For data visualization we used choropleth. A Choropleth Map is a map composed of colored polygons. It is used to represent spatial variations of a quantity. This page documents how to build outline choropleth maps, but you can also build choropleth tile maps using our Mapbox trace types. Below we show how to create Choropleth Maps using either Plotly Express' px.choropleth function or the lower-level go.Choropleth graph object.

```
fig = px.choropleth(df_active_case,
            location = “country”,
            locationmode = “country names”,
            color = “Active”
            hover-name = “country”,
                    animation_frame = “Date”
)
```

C. Build Model (fbprophet)

It’s a tool intended to help you to do time series forecasting at a scale with ease.

FBProphet uses decomposable time series model with 3 main components: seasonal, trends, holidays or events effect and error which are combined into this equation:

$$f(x) = g(x) + s(x) + h(x) + e(t)$$

FBProphet uses time as a regressor and tries to fit several linear and nonlinear functions of time as components. By default, FBProphet will fit the data using a linear model but it can be changed to the nonlinear model (logistics growth) from its arguments.

Before deciding to use FBprophet, there are some data characteristics that should be met:

1) Hourly, daily or weekly observation within at least a few months (minimum 1 month), but 1 year of historical data is much preferred
2) Having strong seasonalities: day of the week and time of the year
3) Important events or holidays that occur must be noted
4) A reasonable number of missing or outlier data
5) Historical trend changes.

```
# Time series analysis for casting
  from fbprophet import Prophet
    #  forecasting for confirmed cases

confirmed=df.groupby(‘Date’)[‘Confirmed’].sum().reset_index()
confirmed.head()    // It will show the confirmed cases from the top of the list with the date
```

We can also forecast per day confirmed cases, deaths, active cases, total recovered. We can also forecast data of India as well as other countries.

D. Prediction (Timeseries)

To make predictions, a dataframe needs to be created with the future dates. In context of FbProphet these are identified as periods are parameters indicating how many days to create. In the methodology adopted, the forecasting consists of 90 days (3 months). The forecasting model to predict for 90 days and in these, we predict data like active cases, confirmed cases and death count country-wise.

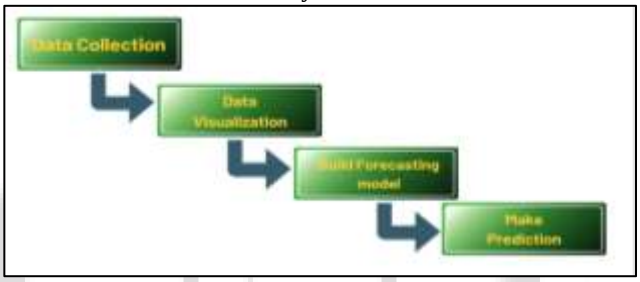


Fig. 2: the flow diagram of proposed forecasting technique

V. EXPERIMENTAL WORKS

A. Data visualisation

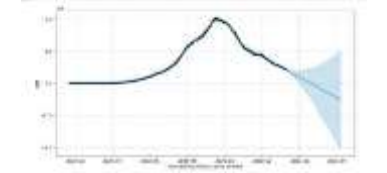
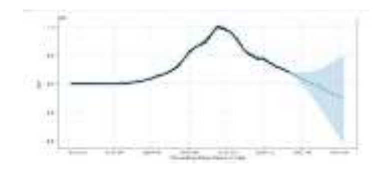
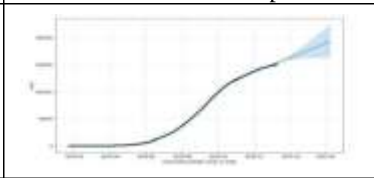
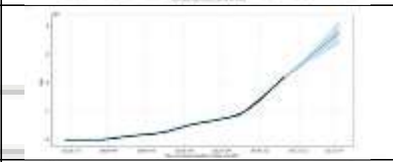
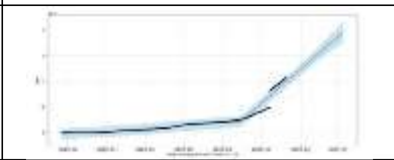
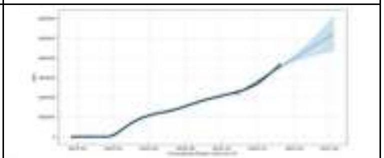
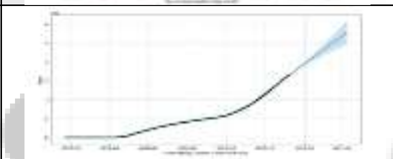
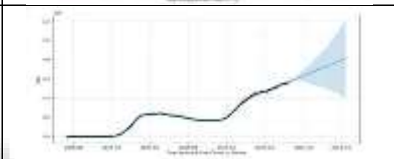
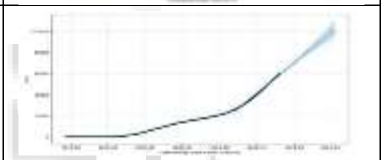
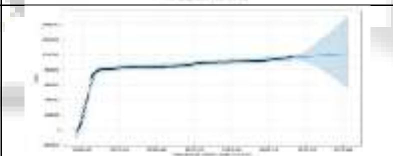
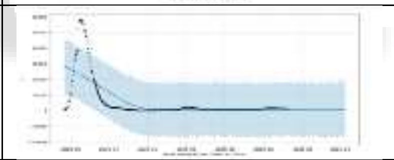
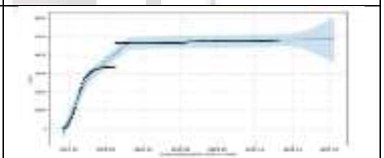
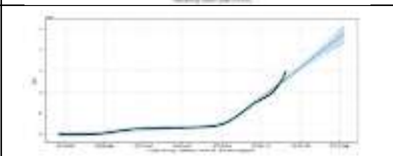
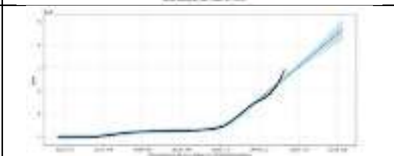
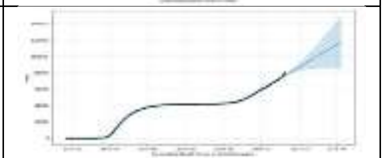
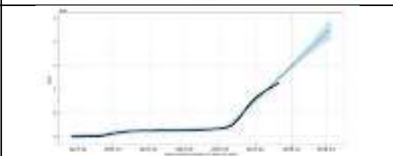
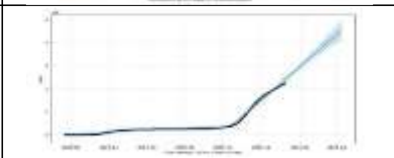
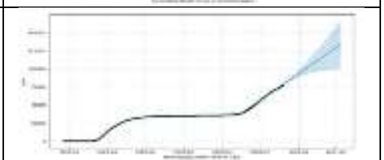
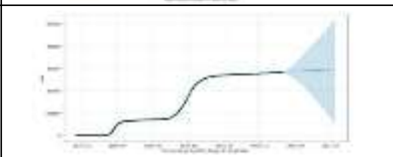
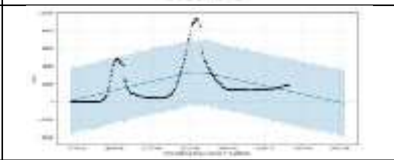
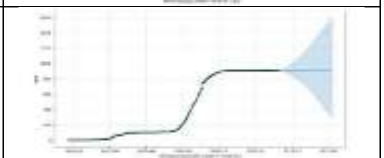
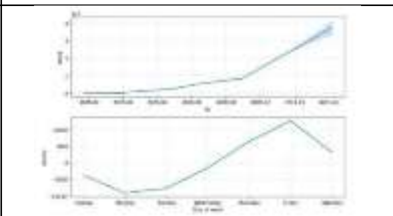
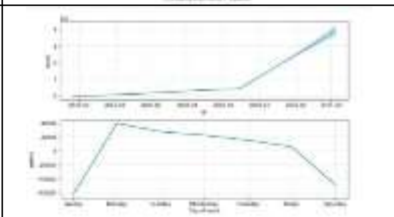
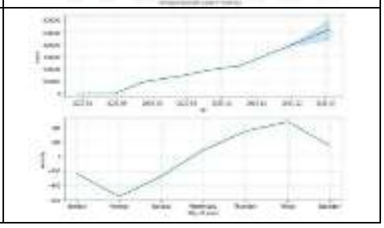
Fig. 3: Confirm Cases

Fig. 4: Confirm Cases

Fig. 5: Active Cases

Fig. 6: Recovery Cases

In this we will show graph of countries with data like active cases, confirmed cases, death count in weekly and monthly basis.

Country	Covid-19 cases graph with forecasted data (Monthly)		
	Confirmed Cases Graph	Active Cases Graph	Death Count Graph
India			
USA			
Russia			
China			
UK			
Italy			
Australia			
USA			

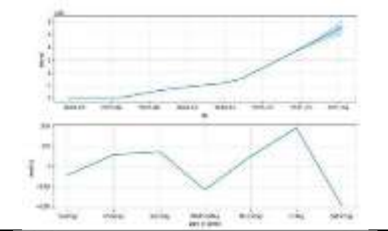
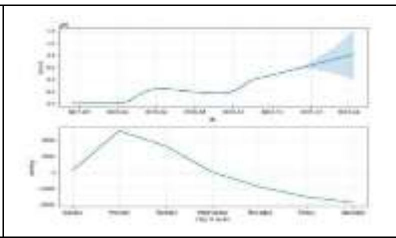
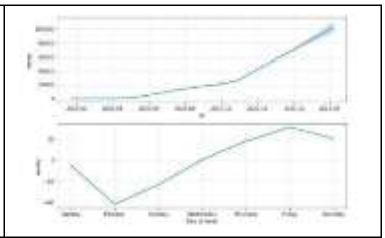
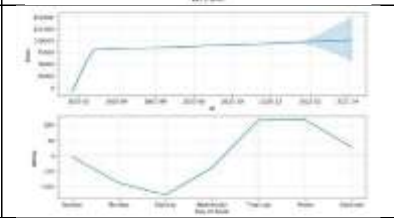
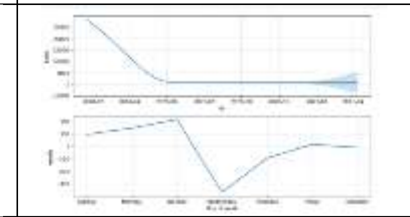
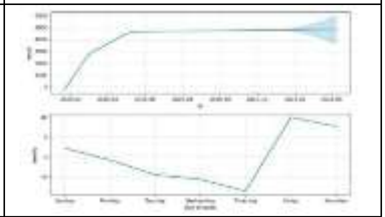
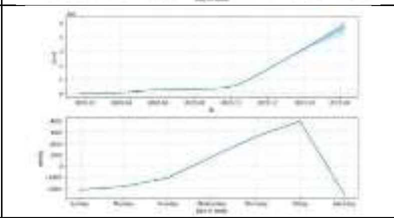
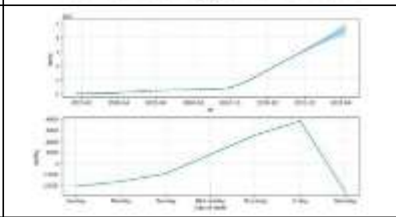
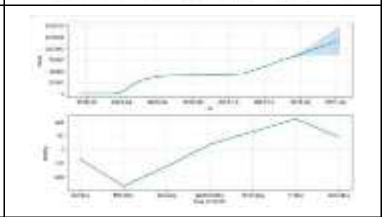
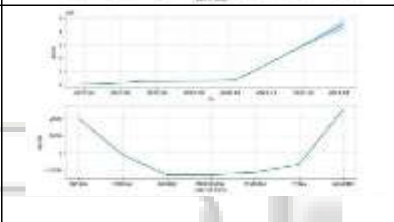
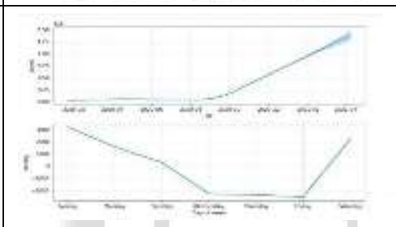
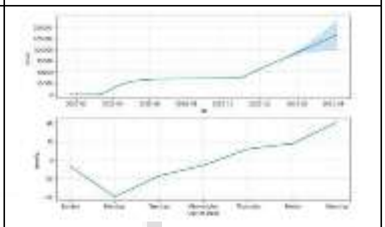
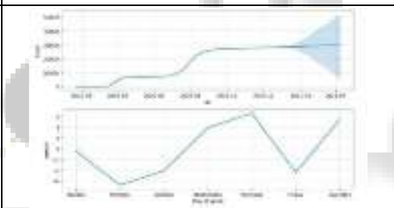
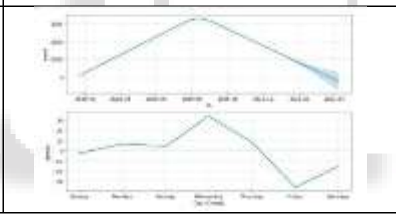
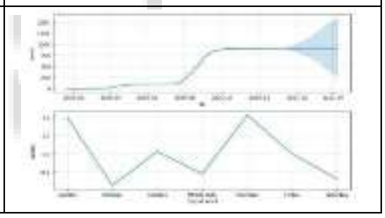
Russia			
China			
UK			
Italy			
Australia			

Table 1:

VI. Conclusion

Information and communication technology help in the decision-making process based on the past data with the data analytics and data mining perspective and using machine learning models we can forecast with ease. The size of data available is huge and gathering information and getting an interesting pattern out of the accumulated data is a challenging task. The prevailing data about confirmed, recovered, active and death across India as well as in other countries for over the time duration helps in predicting and forecasting the near future. We are using data for 90 days forecasting using models like “fbprophet” and machine learning techniques like “choropleth”, so that we know the cases which are increasing in the future. The correctness of the model could be increased by introducing related attributes.

In this sequel, it is important to analyze various forecasting models for COVID-19 in order to empower governments and allied organizations with the more appropriate information possible. An overall comprehensive study on the analysis of COVID-19, it’s forecasting and prediction, are presented in this study. The major contribution of this study is the analysis of several forecasting models available in the literature and their classification. Based on the available forecasting methods, we studied various statistical, analytical, mathematical and medical (symptomatic and asymptomatic) parameters. We hope that providing analysis of forecasting models of COVID-19 to the government and healthcare community, will be more helpful for adapting better intervention policies and explicitly, it will also help to alleviate the alarming effect of this pandemic. The analysis of forecasted data helps the government to create an effective strategy to stop the spreading of COVID-19 cases.

References

[1] Afshin Shoeibi, Marjane Khodatars, Roohallah Alizadehsani, Member, IEEE, Navid Ghassemi, Mahboobeh Jafari, Parisa Moridian, Ali Khadem, Delaram Sadeghi, Sadiq Hussain, Assef Zare, Zahra Alizadeh Sani, Javad Bazeli and Fahime Khozeimeh, Abbas Khosravi, Member, IEEE, Saeid Nahavandi, Fellow, IEEE, U. Rajendra Acharya, Senior Member, IEEE, Peng Shi, Fellow, IEEE, “Automated Detection and Forecasting of COVID-19 using Deep Learning Techniques: A Review”.

[2] S. Tian, W. Hu, L. Niu, H. Liu, H. Xu, and S.-Y. Xiao, “Pulmonary pathology of early phase 2019 novel coronavirus (covid-19) pneumonia in two patients with lung cancer,” Journal of Thoracic Oncology, 2020.
[3] R. Sujath1, Jyotir Moy Chatterjee2, Aboul Ella Hassanien3, “A machine learning forecasting model for COVID-19 pandemic in India".
[4] Forecasting Models for Coronavirus (COVID-19): A Survey of the State-of-the-Art Gitanjali R. Shinde1*, Asmita B. Kalamkar2 , Parikshit N. Mahalle3,4 , Nilanjan Dey5 , Jyotismita Chaki6 , Aboul Ella Hassanien7.
[5] Dianbo Liu, Leonardo Clemente, Canelle Poirie, Xiyu Ding, Matteo Chinazzi, Jessica Davis, Alessandro Vespignani, Mauricio Santillana, “Real-Time Forecasting of the COVID-19 Outbreak in Chinese Provinces: Machine Learning Approach Using Novel Digital Data and Estimates From Mechanistic Models.” JMIR Vol 22, No 8 (2020): August
[6] Cleo Anastassopoulou, Lucia Russo, Athanasios Tsakris, Constantinos Siettos, :Data-based analysis, modelling and forecasting of the COVID-19 outbreak” Plos One https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0230405
[7] K.C. Santosh “COVID-19 Prediction Models and Unexploited Data.” Springer, 2020.
[8] Dianbo Liuy, Leonardo Clementey, Canelle Poiriery, Xiyu Ding, Matteo Chinazzi, Jessica T Davis, Alessandro Vespignani, Mauricio Santillana. “A machine learning methodology for real-time forecasting of the 2019-2020 COVID-19 outbreak using Internet searches, news alerts, and estimates from mechanistic models” arXiv:2004.04019v1
[9] M. S. Razai, K. Doerholt, S. Ladhani, and P. Oakeshott, “Coronavirus disease 2019 (covid-19): a guide for uk gps,” BMJ, vol. 368, 2020.
[10] A. Remuzzi and G. Remuzzi, “Covid-19 and italy: what next?” The Lancet, 2020.
[11] C. Sohrabi, Z. Alsafi, N. ONeill, M. Khan, A. Kerwan, A. Al-Jabir, C. Iosifidis, and R. Agha, “World health organization declares global emergency: A review of the 2019 novel coronavirus (covid-19),” International Journal of Surgery, 2020.
[12] A. E. Hassanien, L. N. Mahdy, K. A. Ezzat, H. H. Elmousalami, and H. A. Ella, “Automatic x-ray covid-19 lung image classification system based on multi-level thresholding and support vector machine,” medRxiv, 2020.
[13] S. Kadry, V. Rajinikanth, S. Rho, N. S. M. Raja, V. S. Rao, and K. P. Thanaraj, “Development of a machine-learning system to classify lung ct scan images into normal/covid-19 class,” arXiv preprint arXiv:2004.13122, 2020.
[14] W. H. Self, D. M. Courtney, C. D. McNaughton, R. G. Wunderink, and J. A. Kline, “High discordance of chest x-ray and computed tomography for detection of pulmonary opacities in ed patients: implications for diagnosing pneumonia,” The American journal of emergency medicine, vol. 31, no. 2, pp. 401–405, 2013.
[15] G. D. Rubin, C. J. Ryerson, L. B. Haramati, N. Sverzellati, J. P. Kanne, S. Raoof, N. W. Schluger, A. Volpi, J.-J. Yim, I. B. Martin et al., “The role of chest imaging in patient management during the covid-19 pandemic: a multinational consensus statement from the fleischner society,” Chest, 2020.
[16] L. Orioli, M. P. Hermans, J.-P. Thissen, D. Maiter, B. Vandeleene, and J.-C. Yombi, “Covid-19 in diabetic patients: related risks and specifics of management,” in Annales D’endocrinologie. Elsevier, 2020.
[17] “https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases,” https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases, 2020.
[18] R. Alizadehsani, M. Behjati, Z. Roshanzamir, S. Hussain, N. Abedini, F. Hasanzadeh, A. Khosravi, A. Shoeibi, M. Roshanzamir, P. Moradnejadet al., “Risk factors prediction, clinical outcomes, and mortality of covid-19 patients,” medRxiv, 2020.
[19] L. J. Kroft, L. van der Velden, I. H. Gir´on, J. J. Roelofs, A. de Roos, and J. Geleijns, “Added value of ultra–low-dose computed tomography, dose equivalent to chest x-ray radiography, for diagnosing chest pathology,” Journal of thoracic imaging, vol. 34, no. 3, p. 179, 2019.
[20] N. Ghassemi, A. Shoeibi, and M. Rouhani, “Deep neural network with generative adversarial networks pre-training for brain tumor classification based on mr images,” Biomedical Signal Processing and Control, vol. 57, p. 101678, 2020.
[21] M. Talo, O. Yildirim, U. B. Baloglu, G. Aydin, and U. R. Acharya, “Convolutional neural networks for multi-class brain disease detection using mri images,” Computerized Medical Imaging and Graphics, vol. 78, p. 101673, 2019.
[22] N. Ghassemi, A. Shoeibi, M. Rouhani, and H. Hosseini-Nejad, “Epileptic seizures detection in eeg signals using tqwt and ensemble learning,” in 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE). IEEE, 2019, pp. 403–408.
[23] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, “Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals,” Computers in biology and medicine, vol.100, pp. 270–278, 2018.
[24] M. Mohammadpoor, A. Shoeibi, H. Zare, and H. Shojaee, “A hierarchical classification method for breast tumor detection,” Iranian Journal of Medical Physics, vol. 13, no. 4, pp. 261–268, 2016.
[25] D. Arefan, A. A. Mohamed, W. A. Berg, M. L. Zuley, J. H. Sumkin, and S. Wu, “Deep learning modeling using normal mammograms for predicting breast cancer risk,” Medical physics, vol. 47, no. 1, pp. 110–118, 2020.
[26] Kaggle Dataset Definition, 2021. [Online]. Available: https://en.wikipedia.org/wiki/Kaggle/. [Accessed: 11-jan-2021]
[27] Choropleth Definition, 2021. [Online]. Available: https://plotly.com/python/choropleth-maps/. [Accessed: 12-jan-2021]
[28] fbprophet Definition, 2021. [Online]. Available: https://medium.com/tokopedia-data/hacking-time-series-forecasting-like-a-pro- with-fbprophet-76f276f0a058/. [Accessed: 12-jan-2021]